GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 15, 2003, 15:44:49 ; Search time 21.9534 Seconds
(without alignments)
2602.858 Million cell updates/sec

Title: US-10-070-464-7
Perfect score: 1933
Sequence: 1 EEDARSAGVATFVLQEEFDR.....RVAIAGAPVTLWIFYDTGYT 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1933	100.0	360	22	AAB47190	Human DPP8 244Glu-
2	1836.5	95.0	882	22	AAB47187	Human DPP8. Homo
3	1836.5	95.0	882	23	AAE24170	Human dipeptidyl p
4	1836.5	95.0	882	23	ABG61591	Human DPPIV relate
5	1836.5	95.0	882	23	AAU74749	Human protease PRT
6	1836.5	95.0	882	23	AAG78415	Amino acid sequenc
7	1836.5	95.0	882	24	ABU07720	Human serine prote
8	1528	79.0	632	22	AAB93565	Human protein sequ
9	1254.5	64.9	724	23	ABB97362	Novel human protei
10	1254.5	64.9	782	23	ABB97361	Novel human protei
11	1220.5	63.1	830	23	AAE24171	Human dipeptidyl p
12	1220.5	63.1	863	23	ABG61592	Human DPPIV relate
13	1220.5	63.1	892	23	ABB98134	Human PMMM Incyte
14	1220.5	63.1	892	23	ABG61602	Human DPRP-2 splic
15	1220.5	63.1	892	23	ABG61604	Human DPRP-2 splic
16	1220.5	63.1	969	23	AAE24168	Human dipeptidyl p
17	1204.5	62.3	847	23	AAE23875	Murine dipeptidyl
18	1204.5	62.3	869	23	AAE24169	Alternative versio
19	1201	62.1	879	23	ABG61607	Human DPRP-2 splic
20	1201	62.1	879	23	ABG61608	Human DPRP-2 splic
21	1199.5	62.1	658	23	ABG61600	Human DPRP-1 splic
22	1199.5	62.1	661	23	ABG61596	Human DPRP-1 splic
23	1199.5	62.1	690	23	ABG61594	Human DPRP-1 splic
24	1190	61.6	580	23	AAE14337	Human protease PRT
25	1183.5	61.2	683	22	AAM40510	Human polypeptide
26	1183.5	61.2	737	22	AAM38724	Human polypeptide
27	1132.5	58.6	832	23	ABG61605	Human DPRP-2 splic
28	1132.5	58.6	832	23	ABG61606	Human DPRP-2 splic
29	1113	57.6	819	23	ABG61609	Human DPRP-2 splic
30	1113	57.6	819	23	ABG61610	Human DPRP-2 splic
31	1086.5	56.2	310	22	AAB47188	Human DPP8 524Phe-
32	1086.5	56.2	310	23	ABB08994	Human dipeptidyl p
33	984	50.9	497	23	ABG64845	Human albumin fusi
34	984	50.9	497	23	AAU96169	Human secreted pro
35	980	50.7	497	23	ABG64844	Human albumin fusi
36	980	50.7	497	23	AAU96192	Human secreted pro
37	944	48.8	508	21	AAB42928	Human ORFX ORF2692
38	931.5	48.2	613	23	ABG61601	Human DPRP-1 splic
39	860.5	44.5	465	22	AAB47189	Human DPP8 318Thr-
40	767	39.7	250	23	ABB99949	Dipeptidyl peptida
41	722	37.4	1042	22	ABB60137	Drosophila melanog
42	722	37.4	1102	22	ABB62029	Drosophila melanog
43	654.5	33.9	720	21	AAB41626	Human ORFX ORF1390
44	505	26.1	358	23	ABG61597	Human DPRP-1 splic
45	403.5	20.9	732	21	AAB18507	Amino acid sequenc

ALIGNMENTS

RESULT 1
AAB47190
ID AAB47190 standard; Protein; 360 AA.
XX
AC AAB47190;
XX
DT 29-JUN-2001 (first entry)
XX
DE Human DPP8 244Glu-341Ile+515Val-776Thr.
XX
KW Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
KW dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;
KW growth hormone deficiency; glucose level; mucosal regeneration;
KW non-insulin dependent diabetes mellitus;glucose intolerance;
KW immunosuppression.
XX
OS Homo sapiens.
XX
PN WO200119866-A1.
XX
PD 22-MAR-2001.
XX
PF 11-SEP-2000; 2000WO-AU01085.
XX
PR 10-SEP-1999; 99AU-0002762.
PR 18-FEB-2000; 2000AU-0005709.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Abbott CA, Gorell MD;
XX
DR WPI; 2001-281520/29.

```
DR   N-PSDB; AAC85697.
XX
PT   New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
PT   substrates, identifying inhibitors of DPP8 catalytic activity which
PT   have therapeutic uses, and for detecting activated T cells -
XX
PS   Claim 5; Page 74-75; 78pp; English.
XX
CC   The sequences given in AAB47188-90 represent fragments of human
CC   dipeptidyl aminopeptidase (DPP8). DPP8 has substrate specificity for
CC   H-Ala-Pro-pNA, H-Gly-Pro-pNA and H-Arg-Pro-pNA. Therefore, it is a
CC   prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable
CC   of hydrolysing the peptide bond C-terminal to Pro in each of these
CC   compounds. DPP8 is homologous with human DPPIV. DPP8 is useful for
CC   cleaving a substrate, and for detecting an activated T cell which
CC   involves measuring the level of DPP8 gene expression in a T cell. The
CC   level of DPP8 expression is detected by detecting the amount of DPP8
CC   RNA in the cell. It is also useful for identifying a molecule capable
CC   of inhibiting the cleavage of the substrate by DPP8. Molecules
CC   identified as inhibiting DPP8 catalytic activity may be useful for
CC   treating diarrhoea, growth hormone deficiency, lowering glucose levels
CC   in non-insulin dependent diabetes mellitus and other disorders
CC   involving glucose intolerance, enhancing mucosal regeneration and
CC   as immunosuppressants.
XX
SQ   Sequence   360 AA;

  Query Match             100.0%;  Score 1933;  DB 22;  Length 360;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-198;
  Matches  360;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIVDEVRRLVYFEGTKDSPLEHHL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIVDEVRRLVYFEGTKDSPLEHHL 120

Qy        121 YVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 YVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCK 180

Qy        181 TKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGP 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGP 240

Qy        241 QVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVE 300

Qy        301 GLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360


RESULT 2
AAB47187
ID   AAB47187 standard; Protein; 882 AA.
XX
AC   AAB47187;
XX
DT   29-JUN-2001  (first entry)
XX
DE   Human DPP8.
XX
KW   Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
KW   dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;
KW   growth hormone deficiency; glucose level; mucosal regeneration;
KW   non-insulin dependent diabetes mellitus;glucose intolerance;
KW   immunosuppression.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Active-site     739
FT                   /note= "Forms part of Ser-Asp-His catalytic triad"
FT   Active-site     817
FT                   /note= "Forms part of Ser-Asp-His catalytic triad"
FT   Active-site     849
FT                   /note= "Forms part of Ser-Asp-His catalytic triad"
XX
PN   WO200119866-A1.
XX
PD   22-MAR-2001.
XX
PF   11-SEP-2000; 2000WO-AU01085.
XX
PR   10-SEP-1999;   99AU-0002762.
PR   18-FEB-2000; 2000AU-0005709.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Abbott CA,  Gorell MD;
XX
DR   WPI; 2001-281520/29.
DR   N-PSDB; AAC85694.
XX
PT   New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
PT   substrates, identifying inhibitors of DPP8 catalytic activity which
PT   have therapeutic uses, and for detecting activated T cells -
XX
PS   Claim 1; Fig 2; 78pp; English.
XX
CC   This sequence represents human dipeptidyl aminopeptidase (DPP8).
CC   DPP8 has substrate specificity for H-Ala-Pro-pNA, H-Gly-Pro-pNA and
CC   H-Arg-Pro-pNA. Therefore, it is a prolyl oligopeptidase and a
CC   dipeptidyl peptidase, because it is capable of hydrolysing the
CC   peptide bond C-terminal to Pro in each of these compounds. DPP8
CC   is homologous with human DPPIV. DPP8 is useful for cleaving a
CC   substrate, and for detecting an activated T cell which involves
CC   measuring the level of DPP8 gene expression in a T cell. The level
CC   of DPP8 expression is detected by detecting the amount of DPP8 RNA
CC   in the cell. It is also useful for identifying a molecule capable
CC   of inhibiting the cleavage of the substrate by DPP8. Molecules
CC   identified as inhibiting DPP8 catalytic activity may be useful for
CC   treating diarrhoea, growth hormone deficiency, lowering glucose levels
CC   in non-insulin dependent diabetes mellitus and other disorders
CC   involving glucose intolerance, enhancing mucosal regeneration and
CC   as immunosuppressants.
XX
SQ   Sequence   882 AA;

  Query Match             95.0%;  Score 1836.5;  DB 22;  Length 882;
  Best Local Similarity   67.5%;  Pred. No. 4.7e-187;
  Matches  360;  Conservative    0;  Mismatches    0;  Indels  173;  Gaps    1;

Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 303

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              ||||||||||||||||||||||||||||||||||||||
Db        304 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 363

Qy         99 ------------------------------------------------------------ 98

Db        364 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 423

Qy         99 ------------------------------------------------------------ 98

Db        424 EETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 483

Qy         99 ----------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                          ||||||||||||||||||||||||||||||
Db        484 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 543
```

```
Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 603

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 663

Qy        248 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 723

Qy        308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 776
```

RESULT 3
AAE24170
ID AAE24170 standard; Protein; 882 AA.
XX
AC AAE24170;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human dipeptidyl peptidase 8 (DPP8) protein.
XX
KW Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
KW autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;
KW graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
KW antiviral; enzyme.
XX
OS Homo sapiens.
XX
PN WO200234900-A1.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-AU01388.
XX
PR 27-OCT-2000; 2000AU-0001078.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Abbott CA, Gorrell MD;
XX
DR WPI; 2002-454646/48.
DR N-PSDB; AAD38956.
XX
PT New dipeptidyl peptidase (DPP) peptides, useful for screening
PT inhibitors of DPP catalytic activity, which may be employed to treat
PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PT rejection and HIV infection -
XX
PS Example; Fig 1; 91pp; English.
XX
CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
CC polynucleotides encoding such proteins. The DPP peptides are useful for
CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
CC rejection and HIV (human immuno deficiency virus) infection. The present
CC sequence is human DPP8 protein.
XX
SQ Sequence 882 AA;

Query Match 95.0%; Score 1836.5; DB 23; Length 882;
Best Local Similarity 67.5%; Pred. No. 4.7e-187;
Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;

```
Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 303

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              ||||||||||||||||||||||||||||||||||||||
Db        304 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 363

Qy         99 ---------------------------------------------------------- 98

Db        364 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 423

Qy         99 ---------------------------------------------------------- 98

Db        424 EETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 483

Qy         99 -------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                             ||||||||||||||||||||||||||||||
Db        484 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 543

Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 603

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 663

Qy        248 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 723

Qy        308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 776
```

RESULT 4
ABG61591
ID ABG61591 standard; Protein; 882 AA.
XX
AC ABG61591;
XX
DT 12-AUG-2002 (first entry)
XX
DE Human DPPIV related serine protease DPRP-1.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder.
XX
OS Homo sapiens.
XX
PN WO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI; 2002-444178/47.
DR N-PSDB; ABK83322.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT -

```
XX
PS   Claim 17; Fig 1; 113pp; English.
XX
CC   The present invention relates to the isolation of novel human serine
CC   proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC   proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC   and nucleic acids encoding them are useful for treating infections
CC   such as fungal, bacterial, protozoan and viral infections, particularly
CC   infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC   pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC   bulimia, Parkinson's disease, acute heart failure, hypotension,
CC   hypertension, urinary retention, osteoporosis, angina pectoris,
CC   stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC   psychotic and neurological disorders (e.g. anxiety, dementia, or
CC   schizophrenia), and dyskinesias. These may also be used in discovering
CC   therapeutic agents for the treatment of reproductive, inflammatory and
CC   metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ   Sequence   882 AA;

  Query Match             95.0%;  Score 1836.5;  DB 23;  Length 882;
  Best Local Similarity   67.5%;  Pred. No. 4.7e-187;
  Matches  360;  Conservative    0;  Mismatches    0;  Indels  173;  Gaps    1;

Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 303

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              ||||||||||||||||||||||||||||||||||||||
Db        304 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 363

Qy         99 ------------------------------------------------------------ 98

Db        364 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 423

Qy         99 ------------------------------------------------------------ 98

Db        424 EETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 483

Qy         99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                            ||||||||||||||||||||||||||||||
Db        484 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 543

Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 603

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 663

Qy        248 RFKGVKYFRLNTLASLGYVVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 RFKGVKYFRLNTLASLGYVVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 723

Qy        308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 776


RESULT 5
AAU74749
ID   AAU74749 standard; Protein; 882 AA.
XX
AC   AAU74749;
XX
DT   09-APR-2002  (first entry)
XX
DE   Human protease PRTS-9 protein sequence.
XX
KW   Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
KW   cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KW   inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KW   cell proliferative disorder; developmental disorder; epilepsy;
KW   Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KW   reproductive disorder; endometriosis.
XX
OS   Homo sapiens.
XX
PN   WO200198468-A2.
XX
PD   27-DEC-2001.
XX
PF   13-JUN-2001; 2001WO-US19178.
XX
PR   16-JUN-2000; 2000US-212336P.
PR   22-JUN-2000; 2000US-213955P.
PR   29-JUN-2000; 2000US-215396P.
PR   07-JUL-2000; 2000US-216821P.
PR   14-JUL-2000; 2000US-218946P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM;
PI   Delegeane AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
PI   Walia NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT;
PI   Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
PI   Kallick DA;
XX
DR   WPI; 2002-090437/12.
DR   N-PSDB; ABK12892.
XX
PT   Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
PT   in the diagnosis, treatment and prevention of gastrointestinal (e.g.
PT   gastritis), cardiovascular (e.g. atherosclerosis) and cell
PT   proliferative (e.g. cancer)disorders -
XX
PS   Claim 1; Page 140-142; 177pp; English.
XX
CC   The present invention relates to twenty one new human proteases,
CC   referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
CC   polypeptides of the invention are useful in the diagnosis, treatment and
CC   prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
CC   Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
CC   myocardial infarction, autoimmune/inflammatory e.g. acquired
CC   immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
CC   proliferative e.g. cancer, developmental e.g. Duchenne and Becker
CC   muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
CC   epilepsy and Alzheimer's disease and reproductive e.g. infertility and
CC   endometriosis disorders. Numerous other examples of each disorder are
CC   given in the specification. The present protein sequence represents
CC   the human protease PRTS-9 protein of the invention.
XX
SQ   Sequence   882 AA;

  Query Match             95.0%;  Score 1836.5;  DB 23;  Length 882;
  Best Local Similarity   67.5%;  Pred. No. 4.7e-187;
  Matches  360;  Conservative    0;  Mismatches    0;  Indels  173;  Gaps    1;

Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 303

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              ||||||||||||||||||||||||||||||||||||||
Db        304 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 363

Qy         99 ------------------------------------------------------------ 98

Db        364 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 423

Qy         99 ------------------------------------------------------------ 98

Db        424 EETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 483
```

```
Qy         99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                            ||||||||||||||||||||||||||||||
Db        484 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 543

Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 603

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 663

Qy        248 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 723

Qy        308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 776
```

RESULT 6
AAG78415
ID AAG78415 standard; Protein; 882 AA.
XX
AC AAG78415;
XX
DT 12-APR-2002 (first entry)
XX
DE Amino acid sequence of 21953 human prolyl oligopeptidase.
XX
KW 21953 prolyl oligopeptidase; antibody; proline; endopeptidase;
KW cancer; cardiovascular disease; autoimmune disease; atopic allergy;
KW neuronal disorder; vascular disorder; prostate disorder; cytostatic;
KW antidiabetic; antiarthritic; antiasthmatic; antiinflammatory;
KW diabetes mellitus; arthritis; multiple sclerosis; asthma;
KW Grave's disease; neuronal disorder; demyelinating disease.
XX
OS Homo sapiens.
XX
PN WO200179473-A2.
XX
PD 25-OCT-2001.
XX
PF 11-APR-2001; 2001WO-US40483.
XX
PR 18-APR-2000; 2000US-197508P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Meyers RA, Williamson M;
XX
DR WPI; 2002-034353/04.
DR N-PSDB; AAH99934.
XX
PT New polypeptides 21953, member of human prolyl oligopeptidase family,
PT useful as diagnostic targets and therapeutic agents for controlling
PT cancer, lymphoma and leukemia -
XX
PS Claim 1; Page 102-103; 121pp; English.
XX
CC This invention relates to an isolated 21953 human prolyl
CC oligopeptidase. Which is cytostatic, antidiabetic, antiarthritic,
CC neuroprotective, antithyroid, dermatological, antipsoriatic,
CC antiasthmatic, ophthalmological, antiinflammatory, nootropic,
CC antiparkinsonian, anticonvulsant, gynaecological, vasotropic,
CC antianginal, cardiant, antiatherosclerotic, anorectic and
CC metabolic in its action. Uses include gene therapy, expression or
CC activity of 21953 protein modulator, it is useful for identifying a
CC compound which binds to it and can be used in preventing, treating
CC or detecting a cellular proliferative or differentiative disorder.
CC The 21953 molecules can act as novel diagnostic targets and therapeutic
CC agents for controlling disorders associated with the aberrant activity
CC or degradation of peptide hormones e.g., disorders associated with cell
CC differentiation and proliferation such as cancer, immune function,
CC reproductive, neurological and cardiovascular function. The 21953
CC molecules are thus useful for treating and preventing cellular
CC proliferative and differentiative disorders, haematopoietic neoplastic
CC disorders, immune disorders such as autoimmune diseases, diabetes
CC mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
CC neuronal disorders, demyelinating diseases, vascular disorders and
CC metabolism or pain disorders. This sequence represents the amino
CC acid sequence of 21953 human prolyl oligopeptidase.
XX
SQ Sequence 882 AA;

Query Match 95.0%; Score 1836.5; DB 23; Length 882;
Best Local Similarity 67.5%; Pred. No. 4.7e-187;
Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;

```
Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 303

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              ||||||||||||||||||||||||||||||||||||||
Db        304 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 363

Qy         99 ------------------------------------------------------------ 98

Db        364 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 423

Qy         99 ------------------------------------------------------------ 98

Db        424 EETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 483

Qy         99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                            ||||||||||||||||||||||||||||||
Db        484 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 543

Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 603

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 663

Qy        248 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 723

Qy        308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 776
```

RESULT 7
ABU07720
ID ABU07720 standard; Protein; 882 AA.
XX
AC ABU07720;
XX
DT 19-MAY-2003 (first entry)
XX
DE Human serine protease HIPHUM46.
XX
KW Human; enzyme; HIPHUM46; serine protease; gene therapy; osteoarthritis;
KW serine protease activity modulation; dipeptidyl peptidase activity;
KW musculoskeletal disease; Hepatitis B virus infection; myotonic dystrophy;
KW Alzheimer's disease; parasupranuclear palsy; Huntington's disease;
KW amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;
KW irritable bowel syndrome; type I diabetes; faecal incontinence; anaemia;
KW haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;
KW colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;
KW multiple sclerosis.

```
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          259..260
FT                   /note= "Paired glutamates of the beta propeller domain"
FT   Active-site     739
FT                   /label= Catalytic_serine_residue
FT   Active-site     817
FT                   /label= Catalytic_aspartate_residue
FT   Active-site     849
FT                   /label= Catalytic_histidine_residue
XX
PN   GB2374869-A.
XX
PD   30-OCT-2002.
XX
PF   22-JAN-2002; 2002GB-0001404.
XX
PR   23-JAN-2001; 2001GB-0001760.
XX
PA   (GLAX ) GLAXO GROUP LTD.
XX
PI   Edbrooke MR,  Lewis AP;
XX
DR   WPI; 2003-150703/15.
DR   N-PSDB; ABX12255.
XX
PT   Identifying modulators of serine protease activity useful for treating
PT   musculoskeletal diseases, by contacting cell expressing a novel serine
PT   protease polypeptide with a compound and monitoring serine protease
PT   activity  -
XX
PS   Claim 10; Page 26-29; 38pp; English.
XX
CC   The invention relates to a method of identifying a substance that
CC   modulates serine protease activity, comprising contacting a cell such as
CC   a neuronal cell, lung cell, intestinal cell or a cell infected with a
CC   virus, expressing a serine protease polypeptide (HIPHUM 46), or its
CC   variant having dipeptidyl peptidase activity, or a serine protease
CC   isolated from the cell with a test substance and monitoring for serine
CC   protease activity. The method is useful for identifying a substance that
CC   modulates serine protease activity. A modulator of the serine protease is
CC   useful in the manufacture of a medicament for treatment or prophylaxis of
CC   a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus
CC   infection, Alzheimer's disease, parasupranuclear palsy, myotonic
CC   dystrophy, Huntington's disease or amyotrophic lateral sclerosis.
CC   Additional disease that may be treated using modulators of the serine
CC   protease include malabsorption syndromes, irritable bowel syndrome, lung
CC   disease, type I diabetes, faecal incontinence, haemorrhoids, proctitis,
CC   rectal polyps, small bowel tumours, colorectal tumours, anaemia,
CC   dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple
CC   sclerosis. The present sequence represents the amino acid sequence of the
CC   human serine protease HIPHUM46.
XX
SQ   Sequence   882 AA;

  Query Match             95.0%;  Score 1836.5;  DB 24;  Length 882;
  Best Local Similarity   67.5%;  Pred. No. 4.7e-187;
  Matches  360;  Conservative    0;  Mismatches    0;  Indels  173;  Gaps    1;

Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 303

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              ||||||||||||||||||||||||||||||||||||
Db        304 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 363

Qy         99 ---------------------------------------------------------- 98

Db        364 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 423

Qy         99 ---------------------------------------------------------- 98

Db        424 EETTDIWININHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 483

Qy         99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                            ||||||||||||||||||||||||||||
Db        484 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 543

Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 603

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 663

Qy        248 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 723

Qy        308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 776
```

```
RESULT 8
AAB93565
ID   AAB93565 standard; Protein; 632 AA.
XX
AC   AAB93565;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:12964.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 12964; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
```

```
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   632 AA;

  Query Match             79.0%;  Score 1528;  DB 22;  Length 632;
  Best Local Similarity   57.8%;  Pred. No. 3.2e-154;
  Matches  308;  Conservative    0;  Mismatches    1;  Indels  224;  Gaps    2;

Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         45 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 104

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              ||||||||||||||||||||||||||||||||||||
Db        105 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 164

Qy         99 ------------------------------------------------------------ 98

Db        165 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 224

Qy         99 ------------------------------------------------------------ 98

Db        225 EETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 284

Qy         99 -------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                             ||||||||||||||||||||||||||||||
Db        285 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 344

Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        345 PGEVTRLTDRSYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 404

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        405 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 464

Qy        248 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307
              |||||||||||||||||||||||||||||||||||||||||||
Db        465 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKM---------------- 507

Qy        308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
                                              ||||||||||||||||||
Db        508 ---------------------------------VAIAGAPVTLWIFYDTGYT 526


RESULT 9
ABB97362
ID   ABB97362 standard; Protein; 724 AA.
XX
AC   ABB97362;
XX
DT   27-JUN-2002  (first entry)
XX
DE   Novel human protein SEQ ID NO: 630.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag.
XX
OS   Homo sapiens.
XX
PN   WO200222660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US26015.
XX
PR   11-SEP-2000; 2000US-0659671.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   N-PSDB; ABN32548.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis -
XX
PS   Example 2; SEQ ID NO 630; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC   stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC   e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a protein of the invention.
XX
SQ   Sequence   724 AA;

  Query Match             64.9%;  Score 1254.5;  DB 23;  Length 724;
  Best Local Similarity   48.8%;  Pred. No. 7.9e-125;
  Matches  260;  Conservative    0;  Mismatches    0;  Indels  273;  Gaps    2;

Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 245

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              ||||||||||||||||||||||||||||||||||||
Db        246 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 305

Qy         99 ------------------------------------------------------------ 98

Db        306 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 365

Qy         99 ------------------------------------------------------------ 98

Db        366 EETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 425

Qy         99 -------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                             ||||||||||||||||||||||||||||||
Db        426 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 485

Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 545

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        546 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQ------ 599

Qy        248 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307
              ||||||||||||||||||||||||||||||||||||||||||||
Db        600 ---------------------------------------------------------- 599

Qy        308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
```

```
                                                  ||||||||||||||||||||
Db        600 ---------------------------------VAIAGAPVTLWIFYDTGYT 618


RESULT 10
ABB97361
ID   ABB97361 standard; Protein; 782 AA.
XX
AC   ABB97361;
XX
DT   27-JUN-2002  (first entry)
XX
DE   Novel human protein SEQ ID NO: 629.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag.
XX
OS   Homo sapiens.
XX
PN   WO200222660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US26015.
XX
PR   11-SEP-2000; 2000US-0659671.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   N-PSDB; ABN32547.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis -
XX
PS   Example 2; SEQ ID NO 629; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC   stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC   e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a protein of the invention.
XX
SQ   Sequence   782 AA;

  Query Match             64.9%;  Score 1254.5;  DB 23;  Length 782;
  Best Local Similarity   48.8%;  Pred. No. 8.9e-125;
  Matches  260;  Conservative    0;  Mismatches    0;  Indels  273;  Gaps    2;

Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 303

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              ||||||||||||||||||||||||||||||||||||
Db        304 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 363

Qy         99 ------------------------------------------------------------ 98

Db        364 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 423

Qy         99 ------------------------------------------------------------ 98

Db        424 EETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 483

Qy         99 ----------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                            ||||||||||||||||||||||||||||||
Db        484 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 543

Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 603

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQ------ 657

Qy        248 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307

Db        658 ------------------------------------------------------------ 657

Qy        308 RYDFIDLLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
                                                  ||||||||||||||||||||
Db        658 ---------------------------------VAIAGAPVTLWIFYDTGYT 676


RESULT 11
AAE24171
ID   AAE24171 standard; Protein; 830 AA.
XX
AC   AAE24171;
XX
DT   23-SEP-2002  (first entry)
XX
DE   Human dipeptidyl peptidase 4 (DPP4)-like 2 protein.
XX
KW   Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
KW   autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;
KW   graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
KW   antiviral; enzyme; DPP-4 like 2 protein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 235
FT                   /note= "Encoded by GAG"
XX
PN   WO200234900-A1.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-AU01388.
XX
PR   27-OCT-2000; 2000AU-0001078.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Abbott CA,  Gorrell MD;
XX
DR   WPI; 2002-454646/48.
DR   N-PSDB; AAD38957.
XX
PT   New dipeptidyl peptidase (DPP) peptides, useful for screening
PT   inhibitors of DPP catalytic activity, which may be employed to treat
PT   e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PT   rejection and HIV infection -
XX
PS   Disclosure; Page 82-86; 91pp; English.
XX
CC   The present invention relates to dipeptidyl peptidase (DPP) proteins and
CC   polynucleotides encoding such proteins. The DPP peptides are useful for
CC   screening inhibitors of DPP catalytic activity. The inhibitors are useful
CC   for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
CC   rejection and HIV (human immuno deficiency virus) infection. The present
CC   sequence is human DPP4-like 2 protein.
XX
```

SQ Sequence 830 AA;

Query Match 63.1%; Score 1220.5; DB 23; Length 830;
Best Local Similarity 44.2%; Pred. No. 4.3e-121;
Matches 236; Conservative 54; Mismatches 69; Indels 175; Gaps 2;

```
Qy          2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
              :| :||||||||:||||||::|||||| |    | | | ||||||| |||||||||| ||
Db        201 DDPKSAGVATFVIQEEFDRFTGYWWCPTASWEGSQGLKTLRILYEEVDESEVEVIHVPSP 260

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
               || |: ||:|||:||: |||:  |::|   |::|:|:
Db        261 ALEERKTDSYRYPRTGSKNPKIALKLAEFQTDSQGKIVSTQEKELVQPFSSLFPKVEYIA 320

Qy         99 ---------------------------------------------------------- 98

Db        321 RAGWTRDGKYAWAMFLDRPQQWLQLVLLPPALFIPSTENEEQRLASARAVPRNVQPYVVY 380

Qy         99 ---------------------------------------------------------- 98

Db        381 EEVTNVWINVHDIFYPFPQSEGEDELCFLRANECKTGFCHLYKVTAVLKSQGYDWSEPFS 440

Qy         99 -------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYV 126
                                              |:|  :||||:||||:||||||||||
Db        441 PGEDEFKCPIKEEIALTSGEWEVLARHGSKIWVNEETKLVYFQGTKDTPLEHHLYVVSYE 500

Qy        127 NPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWA 186
              ||: |||  |:|||| :||: | |:| ||:   | || :|||| |:|||  |   |||
Db        501 AAGEIVRLTTPGFSHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWA 560

Qy        187 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVN 246
              :::::|   ||| ||||| | : :   ||||:|||| ||||||:|||||:||||||||||
Db        561 SMMEAASCPPDYVPPEIFHFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVN 620

Qy        247 NRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLA 306
              | |||:|| |||||||||| ||||| |||| |||:|||| | :|||:||:|||||||::|
Db        621 NSFKGIKYLRLNTLASLGYAVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVA 680

Qy        307 SRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              :| |||| || |||||||||:|||| |: :  :|:||||||||||:|: ||||||
Db        681 EKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVWMAYDTGYT 734
```

RESULT 12
ABG61592
ID ABG61592 standard; Protein; 863 AA.
XX
AC ABG61592;
XX
DT 12-AUG-2002 (first entry)
XX
DE Human DPPIV related serine protease DPRP-2.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder.
XX
OS Homo sapiens.
XX
PN WO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI; 2002-444178/47.
DR N-PSDB; ABK83323.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT -
XX
PS Claim 17; Fig 1; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ Sequence 863 AA;

Query Match 63.1%; Score 1220.5; DB 23; Length 863;
Best Local Similarity 44.2%; Pred. No. 4.6e-121;
Matches 236; Conservative 54; Mismatches 69; Indels 175; Gaps 2;

```
Qy          2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
              :| :||||||||:||||||::|||||| |    | | | ||||||| |||||||||| ||
Db        234 DDPKSAGVATFVIQEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSP 293

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
               || |: ||:|||:||: |||:  |::|   |::|:|:
Db        294 ALEERKTDSYRYPRTGSKNPKIALKLAEFQTDSQGKIVSTQEKELVQPFSSLFPKVEYIA 353

Qy         99 ---------------------------------------------------------- 98

Db        354 RAGWTRDGKYAWAMFLDRPQQWLQLVLLPPALFIPSTENEEQRLASARAVPRNVQPYVVY 413

Qy         99 ---------------------------------------------------------- 98

Db        414 EEVTNVWINVHDIFYPFPQSEGEDELCFLRANECKTGFCHLYKVTAVLKSQGYDWSEPFS 473

Qy         99 -------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYV 126
                                              |:|  :||||:||||:||||||||||
Db        474 PGEDEFKCPIKEEIALTSGEWEVLARHGSKIWVNEETKLVYFQGTKDTPLEHHLYVVSYE 533

Qy        127 NPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWA 186
              ||: |||  |:|||| :||: | |:| ||:   | || :|||| |:|||  |   |||
Db        534 AAGEIVRLTTPGFSHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWA 593

Qy        187 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVN 246
              :::::|   ||| ||||| | : :   ||||:|||| ||||||:|||||:||||||||||
Db        594 SMMEAASCPPDYVPPEIFHFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVN 653

Qy        247 NRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLA 306
              | |||:|| |||||||||| ||||| |||| |||:|||| | :|||:||:|||||||::|
Db        654 NSFKGIKYLRLNTLASLGYAVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVA 713

Qy        307 SRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              :| |||| || |||||||||:|||| |: :  :|:||||||||||:|: ||||||
Db        714 EKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVWMAYDTGYT 767
```

RESULT 13
ABB98134
ID ABB98134 standard; Protein; 892 AA.

XX
AC ABB98134;
XX
DT 17-OCT-2002 (first entry)
XX
DE Human PMMM Incyte ID 7972712CD1.
XX
KW Human; PMMM; protein modification and maintenance molecule;
KW anticonvulsant; neuroprotective; nootropic; cytostatic; antipsoriatic;
KW antiasthmatic; dermatological; antidiabetic; antiparkinsonian;
KW antianaemic; antiinflammatory; antiulcer; antianginal; cardiant;
KW hepatotropic; osteopathic; antiemetic; antipyretic; virucide;
KW antibacterial; fungicide; gastrointestinal; antidiarrheic; laxative;
KW haemostatic; anti-HIV; thrombolytic; anticoagulant; gynaecological;
KW cardiovascular; antiarteriosclerotic; hypotensive; vasotropic;
KW antitumour; antirheumatic; immunosuppressive; antiallergic; antithyroid;
KW nephrotropic; antigout; thyromimetic; antiarthritic; uropathic;
KW ophthalmological; antiparasitic; tranquiliser; vulnerary; keratolytic;
KW auditory; antiseborrheic; antidepressant; neuroleptic; antiinfertility;
KW anthelmintic; protozoacide; Crohn's disease; hypertension;
KW autoimmune; inflammatory; anaemia; cell proliferative; developmental;
KW epithelial; scabies; neurological; Alzheimer's disease; reproductive;
KW ectopic pregnancy; gene therapy; vaccine; disorder;
KW dipeptidyl peptidase 8.
XX
OS Homo sapiens.
XX
PN WO200246383-A2.
XX
PD 13-JUN-2002.
XX
PF 05-DEC-2001; 2001WO-US46964.
XX
PR 08-DEC-2000; 2000US-254399P.
PR 21-DEC-2000; 2000US-257803P.
PR 05-JAN-2001; 2001US-260110P.
PR 19-JAN-2001; 2001US-262851P.
PR 25-JAN-2001; 2001US-264623P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Azimzai Y, Kallick DA, Baughn MR, Griffin JA, Swarnakar A;
PI Lal PG, Walia NK, Hafalia AJA, Gandhi AR, Au-Young J, Elliott VS;
PI Ramkumar J, Thangavelu K, Lu Y, Warren BA, Lu DAM, Lee EA;
PI Tribouley CM, Arvizu C, Delegeane AM, Yao MG, Khan FA;
PI Sanjanwala MM;
XX
DR WPI; 2002-519664/55.
DR N-PSDB; ABQ75955.
XX
PT New isolated Protein Modification and Maintenance polypeptides, useful
PT for diagnosis, and treatment of e.g. gastrointestinal disorders -
XX
PS Claim 1 (a); Page 172-174; 200pp; English.
XX
CC The invention relates to an isolated Protein Modification and Maintenance
CC (PMM) polypeptide. Polypeptides of the invention may be used in the
CC diagnosis, treatment and prevention of disorders associated with
CC decreased expression or activity of PMMM. These include gastrointestinal
CC disorders (e.g. Crohn's disease), cardiovascular disorders (e.g.
CC hypertension), autoimmune/inflammatory disorders (e.g. anaemia), cell
CC proliferative disorders, developmental disorders, epithelial disorders
CC (e.g. scabies), neurological disorders (e.g. Alzheimer's disease)
CC reproductive disorders (e.g. ectopic pregnancy), and in gene therapy or a
CC vaccine for such diseases. They may also be used in the assessment of the
CC effects of exogenous compound on the expression of nucleic acid and amino
CC acid sequences of protein modification and maintenance molecules. The
CC current sequence represents a human PMMM of the invention, which has been
CC found to have homology with human dipeptidyl peptidase 8.
XX
SQ Sequence 892 AA;

Query Match 63.1%; Score 1220.5; DB 23; Length 892;

Best Local Similarity 44.2%; Pred. No. 4.8e-121;
Matches 236; Conservative 54; Mismatches 69; Indels 175; Gaps 2;

```
Qy          2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
              :| :||||||||:||||||::|||||| |    | | | ||||||| ||||||:||| ||
Db        263 DDPKSAGVATFVIQEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSP 322

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
               || |: ||:|||:||: |||:  |::|   |::|:|:
Db        323 ALEERKTDSYRYPRTGSKNPKIALKLAEFQTDSQGKIVSTQEKELVQPFSSLFPKVEYIA 382

Qy         99 ---------------------------------------------------------- 98

Db        383 RAGWTRDGKYAWAMFLDRPQQWLQLVLLPPALFIPSTENEEQRLASARAVPRNVQPYVVY 442

Qy         99 ---------------------------------------------------------- 98

Db        443 EEVTNVWINVHDIFYPFPQSEGEDELCFLRANECKTGFCHLYKVTAVLKSQGYDWSEPFS 502

Qy         99 --------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYV 126
                                               |:|  :||||:||||:||||||||||
Db        503 PGEDEFKCPIKEEIALTSGEWEVLARHGSKIWVNEETKLVYFQGTKDTPLEHHLYVVSYE 562

Qy        127 NPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWA 186
               ||: ||| |:|||| :||: | |:| ||:   | || :|||| |:||| |   |||
Db        563 AAGEIVRLTTPGFSHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWA 622

Qy        187 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVN 246
               :::::|   ||| ||||| | : :   ||||:|||| |||||:|||||:||||||||||
Db        623 SMMEAASCPPDYVPPEIFHFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVN 682

Qy        247 NRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLA 306
               | |||:|| |||||||||| ||||| |||| |||:|||| | :|||:||:|||||||:|
Db        683 NSFKGIKYLRLNTLASLGYAVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVA 742

Qy        307 SRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
               :| |||| || ||||||||:|||| |: :  :|:|||||||||:|: ||||||
Db        743 EKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVWMAYDTGYT 796
```

RESULT 14
ABG61602
ID ABG61602 standard; Protein; 892 AA.
XX
AC ABG61602;
XX
DT 12-AUG-2002 (first entry)
XX
DE Human DPRP-2 splice variant #1.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder.
XX
OS Homo sapiens.
XX
PN WO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI; 2002-444178/47.

```
DR   N-PSDB; ABK83333.
XX
PT   New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT   the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT   viral infections, cancers, allergies, neurological disorders, or pain
PT   -
XX
PS   Disclosure; Page 76-78; 113pp; English.
XX
CC   The present invention relates to the isolation of novel human serine
CC   proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC   proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC   and nucleic acids encoding them are useful for treating infections
CC   such as fungal, bacterial, protozoan and viral infections, particularly
CC   infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC   pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC   bulimia, Parkinson's disease, acute heart failure, hypotension,
CC   hypertension, urinary retention, osteoporosis, angina pectoris,
CC   stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC   psychotic and neurological disorders (e.g. anxiety, dementia, or
CC   schizophrenia), and dyskinesias. These may also be used in discovering
CC   therapeutic agents for the treatment of reproductive, inflammatory and
CC   metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ   Sequence   892 AA;

  Query Match             63.1%;  Score 1220.5;  DB 23;  Length 892;
  Best Local Similarity   44.2%;  Pred. No. 4.8e-121;
  Matches  236;  Conservative   54;  Mismatches   69;  Indels  175;  Gaps    2;

Qy          2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
              :| :|||||||:||||||::|||||  |     | | | |||||||  ||||||:||| ||
Db        263 DDPKSAGVATFVIQEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSP 322

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
               || |: ||:|||:||: |||:  |::|   |::|:|:
Db        323 ALEERKTDSYRYPRTGSKNPKIALKLAEFQTDSQGKIVSTQEKELVQPFSSLFPKVEYIA 382

Qy         99 ---------------------------------------------------------- 98

Db        383 RAGWTRDGKYAWAMFLDRPQQWLQLVLLPPALFIPSTENEEQRLASARAVPRNVQPYVVY 442

Qy         99 ---------------------------------------------------------- 98

Db        443 EEVTNVWINVHDIFYPFPQSEGEDELCFLRANECKTGFCHLYKVTAVLKSQGYDWSEPFS 502

Qy         99 --------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYV 126
                                               |:|  :||||:||||:||||||||||
Db        503 PGEDEFKCPIKEEIALTSGEWEVLARHGSKIWVNEETKLVYFQGTKDTPLEHHLYVVSYE 562

Qy        127 NPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWA 186
               ||: |||  |:|||| :||: | |:| ||:   | || :|||| |:|||  |   |||
Db        563 AAGEIVRLTTPGFSHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWA 622

Qy        187 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVN 246
              :::::|   ||| ||||| | : :   ||||:|||| |||||||:|||||:||||||||
Db        623 SMMEAASCPPDYVPPEIFHFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVN 682

Qy        247 NRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLA 306
              | |||:|| |||||||||| ||||| |||| |||:|||| | :|||:||:|||||||::|
Db        683 NSFKGIKYLRLNTLASLGYAVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVA 742

Qy        307 SRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              :| |||| || |||||||||:|||| |: :  :|:|||||||||:|: ||||||
Db        743 EKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVWMAYDTGYT 796


RESULT 15
ABG61604
ID   ABG61604 standard; Protein; 892 AA.
XX
AC   ABG61604;
XX
DT   12-AUG-2002  (first entry)
XX
DE   Human DPRP-2 splice variant #3.
XX
KW   Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW   DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW   diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW   heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW   ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW   dyskinesia; reproductive disorder; inflammatory disorder;
KW   metabolic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200231134-A2.
XX
PD   18-APR-2002.
XX
PF   12-OCT-2001; 2001WO-US31874.
XX
PR   12-OCT-2000; 2000US-240117P.
XX
PA   (FERR ) FERRING BV.
XX
PI   Qi S,  Akinsanya KO,  Riviere PJ,  Junien J;
XX
DR   WPI; 2002-444178/47.
DR   N-PSDB; ABK83335.
XX
PT   New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT   the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT   viral infections, cancers, allergies, neurological disorders, or pain
PT   -
XX
PS   Disclosure; Page 81-84; 113pp; English.
XX
CC   The present invention relates to the isolation of novel human serine
CC   proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC   proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC   and nucleic acids encoding them are useful for treating infections
CC   such as fungal, bacterial, protozoan and viral infections, particularly
CC   infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC   pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC   bulimia, Parkinson's disease, acute heart failure, hypotension,
CC   hypertension, urinary retention, osteoporosis, angina pectoris,
CC   stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC   psychotic and neurological disorders (e.g. anxiety, dementia, or
CC   schizophrenia), and dyskinesias. These may also be used in discovering
CC   therapeutic agents for the treatment of reproductive, inflammatory and
CC   metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ   Sequence   892 AA;

  Query Match             63.1%;  Score 1220.5;  DB 23;  Length 892;
  Best Local Similarity   44.2%;  Pred. No. 4.8e-121;
  Matches  236;  Conservative   54;  Mismatches   69;  Indels  175;  Gaps    2;

Qy          2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
              :| :|||||||:||||||::|||||  |     | | | |||||||  ||||||:||| ||
Db        263 DDPKSAGVATFVIQEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSP 322

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
               || |: ||:|||:||: |||:  |::|   |::|:|:
Db        323 ALEERKTDSYRYPRTGSKNPKIALKLAEFQTDSQGKIVSTQEKELVQPFSSLFPKVEYIA 382

Qy         99 ---------------------------------------------------------- 98

Db        383 RAGWTRDGKYAWAMFLDRPQQWLQLVLLPPALFIPSTENEEQRLASARAVPRNVQPYVVY 442

Qy         99 ---------------------------------------------------------- 98

Db        443 EEVTNVWINVHDIFYPFPQSEGEDELCFLRANECKTGFCHLYKVTAVLKSQGYDWSEPFS 502
```

```
Qy          99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYV 126
                                             |:|  :||||:||||:||||||||||| 
Db         503 PGEDEFKCPIKEEIALTSGEWEVLARHGSKIWVNEETKLVYFQGTKDTPLEHHLYVVSYE 562

Qy         127 NPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWA 186
                 ||: |||  |:|||| :||: | |:| ||:   | || :|||| |:|||  |   |||
Db         563 AAGEIVRLTTPGFSHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWA 622

Qy         187 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVN 246
               :::::|   ||| ||||| | : :   ||||:|||| ||||||:|||||:|||||||||| 
Db         623 SMMEAASCPPDYVPPEIFHFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVN 682

Qy         247 NRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLA 306
               | |||:|| |||||||||| ||||| |||| |||:|||| | :|||:||:|||||||::| 
Db         683 NSFKGIKYLRLNTLASLGYAVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVA 742

Qy         307 SRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
                :| |||| || ||||||||:|||| |: :  :|:|||||||||:|: ||||||
Db         743 EKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVWMAYDTGYT 796

Search completed: October 15, 2003, 17:10:48
Job time : 23.9534 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:10:56 ; Search time 8.03173 Seconds
(without alignments)
1896.467 Million cell updates/sec

Title: US-10-070-464-7
Perfect score: 1933
Sequence: 1 EEDARSAGVATFVLQEEFDR.....RVAIAGAPVTLWIFYDTGYT 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1086.5	56.2	310	4	US-09-794-236-4	Sequence 4, Appli
2	347.5	18.0	771	4	US-09-462-284-2	Sequence 2, Appli
3	315	16.3	755	5	PCT-US93-07923-3	Sequence 3, Appli
4	315	16.3	759	5	PCT-US93-07923-2	Sequence 2, Appli
5	315	16.3	766	1	US-08-230-491A-3	Sequence 3, Appli
6	315	16.3	766	1	US-08-619-280A-3	Sequence 3, Appli
7	315	16.3	766	2	US-08-940-391-3	Sequence 3, Appli
8	315	16.3	766	4	US-09-794-236-1	Sequence 1, Appli
9	315	16.3	766	4	US-10-002-593-6	Sequence 6, Appli
10	257	13.3	760	1	US-08-230-491A-2	Sequence 2, Appli
11	257	13.3	760	1	US-08-619-280A-2	Sequence 2, Appli
12	257	13.3	760	2	US-08-940-391-2	Sequence 2, Appli
13	204	10.6	632	3	US-09-016-080-1	Sequence 1, Appli
14	196	10.1	657	4	US-09-355-166-1	Sequence 1, Appli
15	166.5	8.6	622	2	US-08-664-646A-2	Sequence 2, Appli
16	166.5	8.6	622	2	US-09-066-285-2	Sequence 2, Appli
17	166.5	8.6	622	3	US-09-261-006-2	Sequence 2, Appli
18	166.5	8.6	622	3	US-08-951-088-2	Sequence 2, Appli
19	166.5	8.6	622	4	US-09-609-566-2	Sequence 2, Appli
20	166.5	8.6	622	4	US-09-609-570-2	Sequence 2, Appli
21	166.5	8.6	622	4	US-09-427-372-2	Sequence 2, Appli
22	166.5	8.6	622	4	US-09-693-554-2	Sequence 2, Appli
23	148.5	7.7	593	5	PCT-US93-07923-11	Sequence 11, Appl
24	147.5	7.6	721	4	US-09-390-234-20	Sequence 20, Appl
25	131.5	6.8	344	4	US-09-724-623-79	Sequence 79, Appl
26	116	6.0	614	4	US-09-252-991A-20060	Sequence 20060, A
27	110.5	5.7	836	4	US-09-491-356C-21	Sequence 21, Appl
28	105	5.4	529	4	US-09-252-991A-24711	Sequence 24711, A
29	104.5	5.4	489	1	US-08-589-893-12	Sequence 12, Appl
30	104.5	5.4	489	2	US-09-020-991-12	Sequence 12, Appl
31	104.5	5.4	489	2	US-09-062-890-12	Sequence 12, Appl
32	104.5	5.4	570	3	US-09-068-960-13	Sequence 13, Appl
33	104	5.4	275	4	US-09-198-452A-181	Sequence 181, App
34	102.5	5.3	489	1	US-08-124-674-2	Sequence 2, Appli
35	102.5	5.3	686	3	US-09-368-169-8	Sequence 8, Appli
36	101.5	5.3	298	4	US-09-355-166-6	Sequence 6, Appli
37	101	5.2	318	4	US-09-355-166-16	Sequence 16, Appl
38	98.5	5.1	444	1	US-08-483-140-28	Sequence 28, Appl
39	98.5	5.1	444	2	US-08-485-938A-32	Sequence 32, Appl
40	98.5	5.1	489	1	US-08-589-893-2	Sequence 2, Appli
41	98.5	5.1	489	1	US-08-589-893-4	Sequence 4, Appli
42	98.5	5.1	489	1	US-08-589-893-6	Sequence 6, Appli
43	98.5	5.1	489	1	US-08-589-893-8	Sequence 8, Appli
44	98.5	5.1	489	1	US-08-589-893-14	Sequence 14, Appl
45	98.5	5.1	489	1	US-08-589-893-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-794-236-4
; Sequence 4, Application US/09794236
; Patent No. 6337069
; GENERAL INFORMATION:
; APPLICANT: Grouzmann, Eric
; APPLICANT: Lacroix, Jean-Silvain
; APPLICANT: Monod, Michel
; TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
; FILE REFERENCE: 81985/276823
; CURRENT APPLICATION NUMBER: US/09/794,236
; CURRENT FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-794-236-4

Query Match 56.2%; Score 1086.5; DB 4; Length 310;
Best Local Similarity 80.6%; Pred. No. 2.2e-106;
Matches 204; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

```
Qy        108 FEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSL 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 FEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSL 60

Qy        168 YKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPG 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPG 120

Qy        228 KKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFK 287
              ||||||||||||||
Db        121 KKYPTVLFIYGGPQ---------------------------------------------- 134

Qy        288 YKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGA 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 ---GQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGA 191

Qy        348 PVTLWIFYDTGYT 360
              |||||||||||||
Db        192 PVTLWIFYDTGYT 204
```

RESULT 2
US-09-462-284-2
; Sequence 2, Application US/09462284
; Patent No. 6309868
; GENERAL INFORMATION:

```
;  APPLICANT:  Nestec S.A.
;  APPLICANT:  Monod, Michel
;  APPLICANT:  Doumas, Agnes
;  APPLICANT:  Affolter, Micheal
;  APPLICANT:  Van Den Broek, Peter
;  TITLE OF INVENTION:  CLONING OF THE
;  TITLE OF INVENTION:  PROLYL-DIPEPTIDYL-PEPTIDASE FROM
;  TITLE OF INVENTION:  ASPERGILLUS ORYZAE
;  FILE REFERENCE:  8265-298
;  CURRENT APPLICATION NUMBER: US/09/462,284
;  CURRENT FILING DATE:  2000-01-03
;  NUMBER OF SEQ ID NOS:  9
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 771
;   TYPE: PRT
;   ORGANISM: Fungus
US-09-462-284-2

  Query Match             18.0%;  Score 347.5;  DB 4;  Length 771;
  Best Local Similarity   25.4%;  Pred. No. 1.1e-27;
  Matches  124;  Conservative   57;  Mismatches  150;  Indels  157;  Gaps   16;

Qy          8 GVATFVLQEEF--DRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLETR 65
              ||  :: :||   |||:  |: |  |          : |   :|: |    |    |
Db        190 GVPDWIYEEEILGDRYA-LWFSPDGE---------YLAYLSFNETGVPTYTVQYYMDNQE 239

Qy         66 RADSF------RYPKTGTANPKVT-----------------FKMSEIMI---------- 91
                | ::      ||||    || ||                 |: ::::|
Db        240 IAPAYPWELKIRYPKVSQTNPTVTLSLLNIASKEVKQAPIDAFESTDLIIGEVAWLTDTH 299

Qy         92 -------------------------------DAEGR----------------------- 96
                                             | :|
Db        300 TTVAAKAFNRVQDQQKVVAVDTASNKATVISDRDGTDGWLDNLLSMKYIGPIKPSDKDAY 359

Qy         97 -----------------------------------IIVDEVRRLVYFEGTKDSPLEHHL 120
                                                  : :|: |:|||:  |:    | ||
Db        360 YIDISDHSGWAHLYLFPVSGGEPIPLTKGDWEVTSILSIDQERQLVYYLSTQHHSTERHL 419

Qy        121 YVVSYVNPGEVTRLTDRGYS--HSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPT 178
              | ||| :   || | |   :   |   | :  ::|  |      |:   || :|
Db        420 YSVSY-STFAVTPLVDDTVAAYWSASFSANSGYYILTYGGPDVPY-QELYTTNS------ 471

Qy        179 CKTKEFWATILDSAGPL---PDYTPPEIFSFE--STTGFTLYGMLYKPHDLQPGKKYPTV 233
                ||    || |:|  |    ||  | |  ||    :| ||  |   |     | |||| :
Db        472 --TKPL-RTITDNAKVLEQIKDYALPNITYFELPLPSGETLNVMQRLPPGFSPDKKYPIL 528

Qy        234 LFIYGGPQVQLVNNRFKGVKY-FRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQ 292
                ||||  | |  |:: : :   : : : | ||   :|||:  :| ||  |   ::|
Db        529 FTPYGGPGAQEVTKRWQALNFKAYVASDSELEYVTWTVDNRGTGFKGRKFRSAVTRQLGL 588

Qy        293 IEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLW 352
              :| :||:   |  |:   :|| | :|| |||:||||:   | : |  | : :  |||: |
Db        589 LEAEDQIYAAQQ-AANIPWIDADHIGIWGWSFGGYLTSKVLEKDSGAFTLGVITAPVSDW 647

Qy        353 IFYDTGYT 360
              |||: ||
Db        648 RFYDSMYT 655


RESULT 3
PCT-US93-07923-3
; Sequence 3, Application PC/TUS9307923
;  GENERAL INFORMATION:
;    APPLICANT:  Morimoto, Chikao
;    APPLICANT:  Schlossman, Stuart F.
;    APPLICANT:  Tanaka, Toshiaki
;    TITLE OF INVENTION:  HUMAN CD26 AND METHODS FOR USE
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;      SOFTWARE:  WordPerfect (Version 5.0)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/07923
;      FILING DATE:  19930819
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/934,162
;      FILING DATE:  21-AUG-1992
;      APPLICATION NUMBER:  07/832,211
;      FILING DATE:  06-FEB-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fraser, Janis K.
;      REGISTRATION NUMBER:  34,819
;      REFERENCE/DOCKET NUMBER:  00530/055002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  755
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
PCT-US93-07923-3

  Query Match             16.3%;  Score 315;  DB 5;  Length 755;
  Best Local Similarity   23.4%;  Pred. No. 3e-24;
  Matches  120;  Conservative   59;  Mismatches  146;  Indels  188;  Gaps   22;

Qy          1 EEDARSAGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH--- 56
              :||    |:  :| :|| |  ||  || |             : | : :::|| :|
Db        179 KEDIIYNGITDWVYEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSF 229

Qy         57 VTSPMLETRRADSFRYPKTGTANPKVTF-----------KMSEIMIDAEGRIIVDEVRR 104
               :  |:  :     ||| |  || | |                : | | |   :::
Db        230 YSDESLQYPKTVRVPYPKAGAVNPTVKFFVVNTDSLSSVTNATSIQITAPASMLIG---- 285

Qy        105 LVYFEGTKDSPLEHHLYVVSYVNPGEVT-----RLTDRGYSHSC----------CI--SQ 147
                        :|:|  |::     ::     |: :      |          |:   |
Db        286 -----------DHYLCDVTWATQERISLQWLRRIQNYSVMDICDYDESSGRWNCLVARQ 333

Qy        148 HCDF----FISKYSNQKNPHCV----SLYKLSSPED-------------DPTCKTKEFWA 186
              | :     :: ::     ||       | ||: | |:             | |  ||  |
Db        334 HIEMSTTGWVGRF-RPSEPHFTLDGNSFYKIISNEEGYRHICYFQIDKKDCTFITKGTWE 392

Qy        187 TI---------------------------------------------------------- 188
              |
Db        393 VIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYTKVTCLSCELNPERCQYYSVSFSK 452

Qy        189 ------LDSAGP-LPDYT----------------------------PPEIFSF--ESTTG 211
                    |  :|| || ||                          | :   |   : |
Db        453 EAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSALDKMLQNVQMPSKKLDFIILNETK 512

Qy        212 FTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVV 268
              |  | |:  ||      |||| :| :| || |       |    ||||    |||   ::|
Db        513 F-WYQMILPPH-FDKSKKYPLLLDVYAGPCSQ------KADTVFRLNWATYLASTENIIV 564

Qy        269 V-IDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGY 327
              | ||| ::| |    |   ::|  |::||:|  :   |:  |:|  |: | |||||||
Db        565 ASFDGRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQF-SKMGFVDNKRIAIWGWSYGGY 623

Qy        328 LSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              :: | |   | :|:  || |||: | :||: ||
Db        624 VTSMVLGSGSGVFKCGIAVAPVSRWEYYDSVYT 656
```

```
RESULT 4
PCT-US93-07923-2
; Sequence 2, Application PC/TUS9307923
;  GENERAL INFORMATION:
;     APPLICANT:  Morimoto, Chikao
;     APPLICANT:  Schlossman, Stuart F.
;     APPLICANT:  Tanaka, Toshiaki
;     TITLE OF INVENTION:  HUMAN CD26 AND METHODS FOR USE
;     NUMBER OF SEQUENCES:  16
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       COMPUTER:  IBM PS/2 Model 50Z or 55SX
;       OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;       SOFTWARE:  WordPerfect (Version 5.0)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/07923
;       FILING DATE:  19930819
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/934,162
;       FILING DATE:  21-AUG-1992
;       APPLICATION NUMBER:  07/832,211
;       FILING DATE:  06-FEB-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fraser, Janis K.
;       REGISTRATION NUMBER:  34,819
;       REFERENCE/DOCKET NUMBER:  00530/055002
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 542-5070
;       TELEFAX:  (617) 542-8906
;       TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  759
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
PCT-US93-07923-2

  Query Match             16.3%;  Score 315;  DB 5;  Length 759;
  Best Local Similarity   23.4%;  Pred. No. 3e-24;
  Matches  120;  Conservative   59;  Mismatches  146;  Indels  188;  Gaps   22;

Qy          1 EEDARSAGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH--- 56
              :||    |:  :| :|| |  ||  || |             : | : :::|| :|
Db        183 KEDIIYNGITDWVYEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSF 233

Qy         57 VTSPMLETRRADSFRYPKTGTANPKVTF------------KMSEIMIDAEGRIIVDEVRR 104
               :  |:  :     ||| |  || | |              : | | |   :::
Db        234 YSDESLQYPKTVRVPYPKAGAVNPTVKFFVVNTDSLSSVTNATSIQITAPASMLIG---- 289

Qy        105 LVYFEGTKDSPLEHHLYVVSYVNPGEVT-----RLTDRGYSHSC----------CI--SQ 147
                        :|:|  |::     ::     |: :      |           |:   |
Db        290 ------------DHYLCDVTWATQERISLQWLRRIQNYSVMDICDYDESSGRWNCLVARQ 337

Qy        148 HCDF----FISKYSNQKNPHCV----SLYKLSSPED------------DPTCKTKEFWA 186
              | :     :: ::     ||      | ||: | |:             | |  ||  |
Db        338 HIEMSTTGWVGRF-RPSEPHFTLDGNSFYKIISNEEGYRHICYFQIDKKDCTFITKGTWE 396

Qy        187 TI---------------------------------------------------------- 188
              |
Db        397 VIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYTKVTCLSCELNPERCQYYSVSFSK 456

Qy        189 ------LDSAGP-LPDYT--------------------------PPEIFSF--ESTTG 211
                    |  :|| || ||                          | :   |   : |

Db        457 EAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSALDKMLQNVQMPSKKLDFIILNETK 516

Qy        212 FTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVV 268
              |  | |:  ||      |||| :| :| || |       |    ||||    |||   ::|
Db        517 F-WYQMILPPH-FDKSKKYPLLLDVYAGPCSQ------KADTVFRLNWATYLASTENIIV 568

Qy        269 V-IDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGY 327
              | ||| ::| |   |   ::|  |::||:|  :   |:  |:|  |: | |||||||
Db        569 ASFDGRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQF-SKMGFVDNKRIAIWGWSYGGY 627

Qy        328 LSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              :: | |   | :|:  || |||: | :||: ||
Db        628 VTSMVLGSGSGVFKCGIAVAPVSRWEYYDSVYT 660


RESULT 5
US-08-230-491A-3
; Sequence 3, Application US/08230491A
; Patent No. 5587299
;  GENERAL INFORMATION:
;     APPLICANT:  Rettig, Wolfgang J.; Scanlan, Matthew J.;
;     APPLICANT:  Garin-Chesa, Pilar; Old, Lloyd J.
;     TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULE CODING FOR
;     TITLE OF INVENTION:  FIBROBLAST ACTIVATION PROTEIN ' AND USES
;     TITLE OF INVENTION:  THEREOF
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FELFE & LYNCH
;       STREET:             805 THIRD AVENUE
;       CITY:  NEW YORK
;       STATE:              NEW YORK
;       COUNTRY:            USA
;       ZIP:                10022
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  DISKETTE 3.5 inch 1.2 MB STORAGE
;       COMPUTER:  IBM PS/2
;       OPERATING SYSTEM:  PC-DOS
;       SOFTWARE:  WORDPERFECT - ASC II
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/230,491A
;       FILING DATE:  20-APRIL-1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Hanson, No. 5587299man D.
;       REGISTRATION NUMBER:  30,946
;       REFERENCE/DOCKET NUMBER:  LUD 330
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 688-9200
;       TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  766 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
US-08-230-491A-3

  Query Match             16.3%;  Score 315;  DB 1;  Length 766;
  Best Local Similarity   23.4%;  Pred. No. 3.1e-24;
  Matches  120;  Conservative   59;  Mismatches  146;  Indels  188;  Gaps   22;

Qy          1 EEDARSAGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH--- 56
              :||    |:  :| :|| |  ||  || |             : | : :::|| :|
Db        190 KEDIIYNGITDWVYEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSF 240

Qy         57 VTSPMLETRRADSFRYPKTGTANPKVTF------------KMSEIMIDAEGRIIVDEVRR 104
               :  |:  :     ||| |  || | |              : | | |   :::
Db        241 YSDESLQYPKTVRVPYPKAGAVNPTVKFFVVNTDSLSSVTNATSIQITAPASMLIG---- 296

Qy        105 LVYFEGTKDSPLEHHLYVVSYVNPGEVT-----RLTDRGYSHSC----------CI--SQ 147
                        :|:|  |::     ::     |: :      |           |:   |
Db        297 ------------DHYLCDVTWATQERISLQWLRRIQNYSVMDICDYDESSGRWNCLVARQ 344

Qy        148 HCDF----FISKYSNQKNPHCV----SLYKLSSPED------------DPTCKTKEFWA 186
```

```
Db        345 HIEMSTTGWVGRF-RPSEPHFTLDGNSFYKIISNEEGYRHICYFQIDKKDCTFITKGTWE 403

Qy        187 TI---------------------------------------------------------- 188
              |
Db        404 VIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYTKVTCLSCELNPERCQYYSVSFSK 463

Qy        189 ------LDSAGP-LPDYT---------------------------PPEIFSF--ESTTG 211
                    |  :|| || ||                          | :   |   : |
Db        464 EAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSALDKMLQNVQMPSKKLDFIILNETK 523

Qy        212 FTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVV 268
              |  | |:  ||     |||| :| :| ||  |     |    ||||    |||   ::|
Db        524 F-WYQMILPPH-FDKSKKYPLLLDVYAGPCSQ------KADTVFRLNWATYLASTENIIV 575

Qy        269 V-IDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGY 327
                | ||| ::| |   |   ::|  |::||:|  :   |:  |:|  |: | |||||||
Db        576 ASFDGRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQF-SKMGFVDNKRIAIWGWSYGGY 634

Qy        328 LSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              :: | |  | :|:  || |||: | :||: ||
Db        635 VTSMVLGSGSGVFKCGIAVAPVSRWEYYDSVYT 667


RESULT 6
US-08-619-280A-3
; Sequence 3, Application US/08619280A
; Patent No. 5767242
;  GENERAL INFORMATION:
;    APPLICANT:  Zimmermann, Rainer; Park, John E.;
;    APPLICANT:  Rettig, Wolfgang; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
;    TITLE OF INVENTION:  ALPHA, AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 2.0 MB  storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/619,280A
;      FILING DATE:  18-MARCH-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/230,491
;      FILING DATE:  20-APRIL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5767242man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5330.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  766 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-619-280A-3

  Query Match             16.3%;  Score 315;  DB 1;  Length 766;
  Best Local Similarity   23.4%;  Pred. No. 3.1e-24;
  Matches  120;  Conservative   59;  Mismatches  146;  Indels  188;  Gaps   22;

Qy          1 EEDARSAGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH--- 56
                :||    |:  :| :|| |  ||  || |            : | : :::|| :|
```

```
Db        190 KEDIIYNGITDWVYEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSF 240

Qy         57 VTSPMLETRRADSFRYPKTGTANPKVTF------------KMSEIMIDAEGRIIVDEVRR 104
                :   |:  :     ||| |  || | |               : | | |   :::
Db        241 YSDESLQYPKTVRVPYPKAGAVNPTVKFFVVNTDSLSSVTNATSIQITAPASMLIG---- 296
                o   c
Qy        105 LVYFEGTKDSPLEHHLYVVSYVNPGEVT-----RLTDRGYSHSC----------CI--SQ 147
                          :|:|  |::     ::     |: :      |          |:   |
Db        297 -----------DHYLCDVTWATQERISLQWLRRIQNYSVMDICDYDESSGRWNCLVARQ 344

Qy        148 HCDF----FISKYSNQKNPHCV----SLYKLSSPED------------DPTCKTKEFWA 186
              | :     :: ::     ||      | ||: | |:             | |  ||  |
Db        345 HIEMSTTGWVGRF-RPSEPHFTLDGNSFYKIISNEEGYRHICYFQIDKKDCTFITKGTWE 403

Qy        187 TI---------------------------------------------------------- 188
              |
Db        404 VIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYTKVTCLSCELNPERCQYYSVSFSK 463

Qy        189 ------LDSAGP-LPDYT---------------------------PPEIFSF--ESTTG 211
                    |  :|| || ||                          | :   |   : |
Db        464 EAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSALDKMLQNVQMPSKKLDFIILNETK 523

Qy        212 FTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVV 268
              |  | |:  ||     |||| :| :| ||  |     |    ||||    |||   ::|
Db        524 F-WYQMILPPH-FDKSKKYPLLLDVYAGPCSQ------KADTVFRLNWATYLASTENIIV 575

Qy        269 V-IDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGY 327
                | ||| ::| |   |   ::|  |::||:|  :   |:  |:|  |: | |||||||
Db        576 ASFDGRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQF-SKMGFVDNKRIAIWGWSYGGY 634

Qy        328 LSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              :: | |  | :|:  || |||: | :||: ||
Db        635 VTSMVLGSGSGVFKCGIAVAPVSRWEYYDSVYT 667


RESULT 7
US-08-940-391-3
; Sequence 3, Application US/08940391
; Patent No. 5965373
;  GENERAL INFORMATION:
;    APPLICANT:  Zimmermann, Rainer; Park, John E.;
;    APPLICANT:  Rettig, Wolfgang; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED DIMERIC FIBROBLAST ACTIVATION
;    TITLE OF INVENTION:  PROTEIN ALPHA, AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 2.0 MB  storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/940,391
;      FILING DATE:  01-OCT-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/619,280
;      FILING DATE:  18-MARCH-1996
;      APPLICATION NUMBER:  08/230,491
;      FILING DATE:  20-APRIL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5965373man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5330.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
```

```
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  766 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-940-391-3

  Query Match             16.3%;  Score 315;  DB 2;  Length 766;
  Best Local Similarity   23.4%;  Pred. No. 3.1e-24;
  Matches  120;  Conservative   59;  Mismatches  146;  Indels  188;  Gaps   22;

Qy          1 EEDARSAGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH--- 56
              :||    |:  :| :|| |  ||  || |              : | : :::|| :|
Db        190 KEDIIYNGITDWVYEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSF 240

Qy         57 VTSPMLETRRADSFRYPKTGTANPKVTF------------KMSEIMIDAEGRIIVDEVRR 104
               :   |:  :     ||| |  || | |              : | | |   :::
Db        241 YSDESLQYPKTVRVPYPKAGAVNPTVKFFVVNTDSLSSVTNATSIQITAPASMLIG---- 296

Qy        105 LVYFEGTKDSPLEHHLYVVSYVNPGEVT-----RLTDRGYSHSC----------CI--SQ 147
                      :|:|  |::      ::      |: :       |           |:   |
Db        297 ------------DHYLCDVTWATQERISLQWLRRIQNYSVMDICDYDESSGRWNCLVARQ 344

Qy        148 HCDF----FISKYSNQKNPHCV----SLYKLSSPED------------DPTCKTKEFWA 186
              | :      :: ::     ||      | ||: | |:                | |  ||  |
Db        345 HIEMSTTGWVGRF-RPSEPHFTLDGNSFYKIISNEEGYRHICYFQIDKKDCTFITKGTWE 403

Qy        187 TI---------------------------------------------------------- 188
              |
Db        404 VIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYTKVTCLSCELNPERCQYYSVSFSK 463

Qy        189 ------LDSAGP-LPDYT--------------------------PPEIFSF--ESTTG 211
                    |  :|| || ||                          | :   |   : |
Db        464 EAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSALDKMLQNVQMPSKKLDFIILNETK 523

Qy        212 FTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVV 268
              |  | |:  ||      |||| :| :| ||  |       |    ||||    |||   ::|
Db        524 F-WYQMILPPH-FDKSKKYPLLLDVYAGPCSQ------KADTVFRLNWATYLASTENIIV 575

Qy        269 V-IDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGY 327
                | ||| ::| |   |   ::|  |::||:|  :   |:  |:|  |: | |||||||
Db        576 ASFDGRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQF-SKMGFVDNKRIAIWGWSYGGY 634

Qy        328 LSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              :: | |   | :|:  || |||: | :||: ||
Db        635 VTSMVLGSGSGVFKCGIAVAPVSRWEYYDSVYT 667


RESULT 8
US-09-794-236-1
; Sequence 1, Application US/09794236
; Patent No. 6337069
; GENERAL INFORMATION:
;  APPLICANT: Grouzmann, Eric
;  APPLICANT:  Lacroix, Jean-Silvain
;  APPLICANT:  Monod, Michel
;  TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
;  FILE REFERENCE: 81985/276823
;  CURRENT APPLICATION NUMBER: US/09/794,236
;  CURRENT FILING DATE:  2001-02-28
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 766
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-794-236-1

  Query Match             16.3%;  Score 315;  DB 4;  Length 766;
  Best Local Similarity   23.4%;  Pred. No. 3.1e-24;
  Matches  120;  Conservative   59;  Mismatches  146;  Indels  188;  Gaps   22;

Qy          1 EEDARSAGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH--- 56
              :||    |:  :| :|| |  ||  || |              : | : :::|| :|
Db        190 KEDIIYNGITDWVYEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSF 240

Qy         57 VTSPMLETRRADSFRYPKTGTANPKVTF------------KMSEIMIDAEGRIIVDEVRR 104
               :   |:  :     ||| |  || | |              : | | |   :::
Db        241 YSDESLQYPKTVRVPYPKAGAVNPTVKFFVVNTDSLSSVTNATSIQITAPASMLIG---- 296

Qy        105 LVYFEGTKDSPLEHHLYVVSYVNPGEVT-----RLTDRGYSHSC----------CI--SQ 147
                      :|:|  |::      ::      |: :       |           |:   |
Db        297 ------------DHYLCDVTWATQERISLQWLRRIQNYSVMDICDYDESSGRWNCLVARQ 344

Qy        148 HCDF----FISKYSNQKNPHCV----SLYKLSSPED------------DPTCKTKEFWA 186
              | :      :: ::     ||      | ||: | |:                | |  ||  |
Db        345 HIEMSTTGWVGRF-RPSEPHFTLDGNSFYKIISNEEGYRHICYFQIDKKDCTFITKGTWE 403

Qy        187 TI---------------------------------------------------------- 188
              |
Db        404 VIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYTKVTCLSCELNPERCQYYSVSFSK 463

Qy        189 ------LDSAGP-LPDYT--------------------------PPEIFSF--ESTTG 211
                    |  :|| || ||                          | :   |   : |
Db        464 EAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSALDKMLQNVQMPSKKLDFIILNETK 523

Qy        212 FTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVV 268
              |  | |:  ||      |||| :| :| ||  |       |    ||||    |||   ::|
Db        524 F-WYQMILPPH-FDKSKKYPLLLDVYAGPCSQ------KADTVFRLNWATYLASTENIIV 575

Qy        269 V-IDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGY 327
                | ||| ::| |   |   ::|  |::||:|  :   |:  |:|  |: | |||||||
Db        576 ASFDGRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQF-SKMGFVDNKRIAIWGWSYGGY 634

Qy        328 LSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              :: | |   | :|:  || |||: | :||: ||
Db        635 VTSMVLGSGSGVFKCGIAVAPVSRWEYYDSVYT 667


RESULT 9
US-10-002-593-6
; Sequence 6, Application US/10002593
; Patent No. 6586198
; GENERAL INFORMATION:
;  APPLICANT: Vanderbilt University
;  APPLICANT:  Brown, Nancy J.
;  TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOTENSIN CONVER
;  TITLE OF INVENTION:  INHIBITOR AND VASOPEPTIDASE INHIBITOR ASSOCIATED ANGIOEDEMA
;  FILE REFERENCE: Atty Docket No. 6586198 1242/48/2
;  CURRENT APPLICATION NUMBER: US/10/002,593
;  CURRENT FILING DATE:  2001-10-31
;  PRIOR APPLICATION NUMBER: 60/244,524
;  PRIOR FILING DATE: 2000-10-31
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 766
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-002-593-6

  Query Match             16.3%;  Score 315;  DB 4;  Length 766;
  Best Local Similarity   23.4%;  Pred. No. 3.1e-24;
  Matches  120;  Conservative   59;  Mismatches  146;  Indels  188;  Gaps   22;

Qy          1 EEDARSAGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH--- 56
              :||    |:  :| :|| |  ||  || |              : | : :::|| :|
Db        190 KEDIIYNGITDWVYEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSF 240

Qy         57 VTSPMLETRRADSFRYPKTGTANPKVTF------------KMSEIMIDAEGRIIVDEVRR 104
               :   |:  :     ||| |  || | |              : | | |   :::
Db        241 YSDESLQYPKTVRVPYPKAGAVNPTVKFFVVNTDSLSSVTNATSIQITAPASMLIG---- 296

Qy        105 LVYFEGTKDSPLEHHLYVVSYVNPGEVT-----RLTDRGYSHSC----------CI--SQ 147
                      :|:|  |::      ::      |: :       |           |:   |
Db        297 ------------DHYLCDVTWATQERISLQWLRRIQNYSVMDICDYDESSGRWNCLVARQ 344
```

```
Qy        148 HCDF----FISKYSNQKNPHCV----SLYKLSSPED------------DPTCKTKEFWA 186
              |  :     :: ::     ||       | ||: | |:             | |  ||  |
Db        345 HIEMSTTGWVGRF-RPSEPHFTLDGNSFYKIISNEEGYRHICYFQIDKKDCTFITKGTWE 403

Qy        187 TI---------------------------------------------------------- 188
              |
Db        404 VIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLIDYTKVTCLSCELNPERCQYYSVSFSK 463

Qy        189 ------LDSAGP-LPDYT--------------------------PPEIFSF--ESTTG 211
                    |  :|| || ||                         | :   |   : |
Db        464 EAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSALDKMLQNVQMPSKKLDFIILNETK 523

Qy        212 FTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVV 268
              |  | |:  ||     ||||  :| :| ||  |       |    ||||.  |||   ::|
Db        524 F-WYQMILPPH-FDKSKKYPLLLDVYAGPCSQ------KADTVFRLNWATYLASTENIIV 575

Qy        269 V-IDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGY 327
              | ||| ::| |   |   ::|  |::||:|  :   |:  |:|  |: | |||||||
Db        576 ASFDGRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQF-SKMGFVDNKRIAIWGWSYGGY 634

Qy        328 LSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              :: | |   | :|:  || |||: | :||: ||
Db        635 VTSMVLGSGSGVFKCGIAVAPVSRWEYYDSVYT 667


RESULT 10
US-08-230-491A-2
; Sequence 2, Application US/08230491A
; Patent No. 5587299
;  GENERAL INFORMATION:
;    APPLICANT:  Rettig, Wolfgang J.; Scanlan, Matthew J.;
;    APPLICANT:  Garin-Chesa, Pilar; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULE CODING FOR
;    TITLE OF INVENTION:  FIBROBLAST ACTIVATION PROTEIN ' AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FELFE & LYNCH
;      STREET:  805 THIRD AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE 3.5 inch 1.2 MB STORAGE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  WORDPERFECT - ASC II
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/230,491A
;      FILING DATE:  20-APRIL-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5587299man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 330
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  760 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-230-491A-2

  Query Match             13.3%;  Score 257;  DB 1;  Length 760;
  Best Local Similarity   22.7%;  Pred. No. 4.1e-18;
  Matches  110;  Conservative   68;  Mismatches  158;  Indels  148;  Gaps   20;

Qy          8 GVATFVLQEEF--DRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLET- 64
              |:  :| :||    :|:  || |        || |   | | :: :: :|  :    |
```

```
Db        195 GIPDWVYEEEMLPTKYA-LWWSP-------NGKFL--AYAEFNDKDIPVIAYSYYGDEQY 244

Qy         65 RRADSFRYPKTGTANPKVTFKM-----------SEIMIDA----------------EGRI 97
               |  :  ||| |  || |   :            |: : |                : |:
Db        245 PRTINIPYPKAGAKNPVVRIFIIDTTYPAYVGPQEVPVPAMIASSDYYFSWLTWVTDERV 304

Qy         98 IVDEVRRLVYFE-----------GTKDSP-LEHHLY---------------VVSYVNPGE 130
              :  ::|:                | | |  : |:                 | ||
Db        305 CLQWLKRVQNVSVLSICDFREDWQTWDCPKTQEHIEESRTGWAGGFFVSRPVFSYDAISY 364

Qy        131 VTRLTDR-GYSHSCCI----------------------SQHCDFFIS----KYSNQKNP 162
              :|: || |  |                         :|   |: |    :|  ::|
Db        365 YKIFSDKDGYKHIHYIKDTVENAIQITSGKWEAINIFRVTQDSLFYSSNEFEEYPGRRNI 424

Qy        163 HCVSLYKLSSPEDDPTCKTKE-----FWATILDSA--------GP-LPDYT--------- 199
              : :|:      :   ||  ::     : |:  | |        || :|  |
Db        425 YRISIGSYPPSKKCVTCHLRKERCQYYTASFSDYAKYYALVCYGPGIPISTLHDGRTDQE 484

Qy        200 -------------------PPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGP 240
                              | |          ||:  :  |      |||| :: :||||
Db        485 IKILEENKELENALKNIQLPKEEIKKLEVDEITLWYKMILPPQFDRSKKYPLLIQVYGGP 544

Qy        241 QVQLVNNRFKGVKYFRLNTLASL----GYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEID 296
              | | :       | :| :: |    | |: ::| ||:  :| | |   |:|  |::
Db        545 CSQSVRS------VFAVNWISYLASKEGMVIALVDGRGTAFQGDKLLYAVYRKLGVYEVE 598

Qy        297 DQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYD 356
              ||:  ::       |||  |: | |||||||:|  :||   : :|:  || |||: | :|
Db        599 DQITAVRKFI-EMGFIDEKRIAIWGWSYGGYVSSLALASGTGLFKCGIAVAPVSSWEYYA 657

Qy        357 TGYT 360
              : ||
Db        658 SVYT 661


RESULT 11
US-08-619-280A-2
; Sequence 2, Application US/08619280A
; Patent No. 5767242
;  GENERAL INFORMATION:
;    APPLICANT:  Zimmermann, Rainer; Park, John E.;
;    APPLICANT:  Rettig, Wolfgang; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
;    TITLE OF INVENTION:  ALPHA, AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 2.0 MB  storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/619,280A
;      FILING DATE:  18-MARCH-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/230,491
;      FILING DATE:  20-APRIL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5767242man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5330.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  760 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-619-280A-2

  Query Match             13.3%;  Score 257;  DB 1;  Length 760;
  Best Local Similarity   22.7%;  Pred. No. 4.1e-18;
  Matches  110;  Conservative   68;  Mismatches  158;  Indels  148;  Gaps   20;

Qy          8 GVATFVLQEEF--DRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLET- 64
              |:  :| :||    :|:  || |         || |   | | :: :: :|  :    |
Db        195 GIPDWVYEEEMLPTKYA-LWWSP-------NGKFL--AYAEFNDKDIPVIAYSYYGDEQY 244

Qy         65 RRADSFRYPKTGTANPKVTFKM----------SEIMIDA---------------EGRI 97
               |  :  ||| |  || |   :           |: : |               : |:
Db        245 PRTINIPYPKAGAKNPVVRIFIIDTTYPAYVGPQEVPVPAMIASSDYYFSWLTWVTDERV 304

Qy         98 IVDEVRRLVYFE----------GTKDSP-LEHHLY--------------VVSYVNPGE 130
               :  ::|:                | | |  : |:               | ||
Db        305 CLQWLKRVQNVSVLSICDFREDWQTWDCPKTQEHIEESRTGWAGGFFVSRPVFSYDAISY 364

Qy        131 VTRLTDR-GYSHSCCI----------------------SQHCDFFIS----KYSNQKNP 162
              :|: || |   |                        :|   |: |    :|  ::|
Db        365 YKIFSDKDGYKHIHYIKDTVENAIQITSGKWEAINIFRVTQDSLFYSSNEFEEYPGRRNI 424

Qy        163 HCVSLYKLSSPEDDPTCKTKE-----FWATILDSA--------GP-LPDYT--------- 199
              : :|:      :   ||  ::     : |:  | |        || :|  |
Db        425 YRISIGSYPPSKKCVTCHLRKERCQYYTASFSDYAKYYALVCYGPGIPISTLHDGRTDQE 484

Qy        200 ------------------PPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGP 240
                             | |          ||:  :  |      |||| :: :||||
Db        485 IKILEENKELENALKNIQLPKEEIKKLEVDEITLWYKMILPPQFDRSKKYPLLIQVYGGP 544

Qy        241 QVQLVNNRFKGVKYFRLNTLASL----GYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEID 296
              | | :       | :|  :: |     | |: ::| ||:  :| |    |    |:|  |::
Db        545 CSQSVRS------VFAVNWISYLASKEGMVIALVDGRGTAFQGDKLLYAVYRKLGVYEVE 598

Qy        297 DQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYD 356
              ||:  ::       |||  |: | |||||||:| :||   : :|:  || |||: | :|
Db        599 DQITAVRKFI-EMGFIDEKRIAIWGWSYGGYVSSLALASGTGLFKCGIAVAPVSSWEYYA 657

Qy        357 TGYT 360
              : ||
Db        658 SVYT 661


RESULT 12
US-08-940-391-2
; Sequence 2, Application US/08940391
; Patent No. 5965373
;  GENERAL INFORMATION:
;    APPLICANT:  Zimmermann, Rainer; Park, John E.;
;    APPLICANT:  Rettig, Wolfgang; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED DIMERIC FIBROBLAST ACTIVATION
;    TITLE OF INVENTION:  PROTEIN ALPHA, AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 2.0 MB  storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/940,391
;      FILING DATE:  01-OCT-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/619,280
;      FILING DATE:  18-MARCH-1996
;      APPLICATION NUMBER:  08/230,491
;      FILING DATE:  20-APRIL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5965373man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5330.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  760 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-940-391-2

  Query Match             13.3%;  Score 257;  DB 2;  Length 760;
  Best Local Similarity   22.7%;  Pred. No. 4.1e-18;
  Matches  110;  Conservative   68;  Mismatches  158;  Indels  148;  Gaps   20;

Qy          8 GVATFVLQEEF--DRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLET- 64
              |:  :| :||    :|:  || |         || |   | | :: :: :|  :    |
Db        195 GIPDWVYEEEMLPTKYA-LWWSP-------NGKFL--AYAEFNDKDIPVIAYSYYGDEQY 244

Qy         65 RRADSFRYPKTGTANPKVTFKM----------SEIMIDA---------------EGRI 97
               |  :  ||| |  || |   :           |: : |               : |:
Db        245 PRTINIPYPKAGAKNPVVRIFIIDTTYPAYVGPQEVPVPAMIASSDYYFSWLTWVTDERV 304

Qy         98 IVDEVRRLVYFE----------GTKDSP-LEHHLY--------------VVSYVNPGE 130
               :  ::|:                | | |  : |:               | ||
Db        305 CLQWLKRVQNVSVLSICDFREDWQTWDCPKTQEHIEESRTGWAGGFFVSRPVFSYDAISY 364

Qy        131 VTRLTDR-GYSHSCCI----------------------SQHCDFFIS----KYSNQKNP 162
              :|: || |   |                        :|   |: |    :|  ::|
Db        365 YKIFSDKDGYKHIHYIKDTVENAIQITSGKWEAINIFRVTQDSLFYSSNEFEEYPGRRNI 424

Qy        163 HCVSLYKLSSPEDDPTCKTKE-----FWATILDSA--------GP-LPDYT--------- 199
              : :|:      :   ||  ::     : |:  | |        || :|  |
Db        425 YRISIGSYPPSKKCVTCHLRKERCQYYTASFSDYAKYYALVCYGPGIPISTLHDGRTDQE 484

Qy        200 ------------------PPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGP 240
                             | |          ||:  :  |      |||| :: :||||
Db        485 IKILEENKELENALKNIQLPKEEIKKLEVDEITLWYKMILPPQFDRSKKYPLLIQVYGGP 544

Qy        241 QVQLVNNRFKGVKYFRLNTLASL----GYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEID 296
              | | :       | :|  :: |     | |: ::| ||:  :| |    |    |:|  |::
Db        545 CSQSVRS------VFAVNWISYLASKEGMVIALVDGRGTAFQGDKLLYAVYRKLGVYEVE 598

Qy        297 DQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYD 356
              ||:  ::       |||  |: | |||||||:| :||   : :|:  || |||: | :|
Db        599 DQITAVRKFI-EMGFIDEKRIAIWGWSYGGYVSSLALASGTGLFKCGIAVAPVSSWEYYA 657

Qy        357 TGYT 360
              : ||
Db        658 SVYT 661


RESULT 13
US-09-016-080-1
; Sequence 1, Application US/09016080
; Patent No. 6133012
; GENERAL INFORMATION:
;  APPLICANT: Ishikawa, Kaxuhiko
;  APPLICANT: Matsui, Ikuo
;  APPLICANT: Ishida, Hiroyasu
;  APPLICANT: Kosugi, Yoshisugu
;  APPLICANT: Higuchi, Katsuhiko
;  TITLE OF INVENTION: THERMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING
;  TITLE OF INVENTION: THE SAME
;  FILE REFERENCE: 07898/022001
;  CURRENT APPLICATION NUMBER: US/09/016,080
```

```
;  CURRENT FILING DATE: 1998-01-30
;  EARLIER APPLICATION NUMBER: JAPAN 18381/1997
;  EARLIER FILING DATE: 1997-01-31
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 632
;   TYPE: PRT
;   ORGANISM: Pyrococcus horikoshii
US-09-016-080-1

  Query Match             10.6%;  Score 204;  DB 3;  Length 632;
  Best Local Similarity   26.3%;  Pred. No. 1.2e-12;
  Matches   83;  Conservative   53;  Mismatches  123;  Indels   56;  Gaps   17;

Qy          67 ADSFRY-PKTGTANPKVTFKM-----SEIMIDAEGRIIVDEVRRLVYFEG-----TKDSP 115
               |  : | ||||    |:|  :     : :  |  |     :   ||| ||       | |
Db         250 AHIYHYDPKTGELK-KLTKDLDRNAYNSLNSDVRG----SQRAELVYKEGWIYYVATDGP 304

Qy         116 LEHHLYVVSYVNPGEVTRLT--DRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSP 173
                 :|: |:    |:: |:   ||             | |  |        |:  :|
Db         305 -RANLFRVNL--DGKIERVIGGDRSVE---------SFDIGDYIAFTAQDAVTPTELYIY 352

Qy         174 EDDPTCKTKEF--WATILDSAGPLPDYT--PPEIFSFESTTGFTLYGMLYKPHDLQPGKK 229
               |    |  :|  |          :  ||   || |  ::: |  :   : || : : |||
Db         353 RDGKEKKVTDFNKW---------IKGYTLSKPEHFKVKASDGVEIDAWVMKPVNFRKGKK 403

Qy         230 YPTVLFIYGGPQVQLVNNRFKGVKYF-RLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKY 288
               || :| |:|||:          |  :    : | | |:||:  : |||   | :|    :
Db         404 YPAILEIHGGPKTAY------GYAFMHEFHVLTSKGFVVIFSNPRGSDGYGEEF-ADIRG 456

Qy         289 KMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAP 348
                 |: :  | :| :      |:|||| :|:|: | ||||:::   ::  :: |: |:
Db         457 HYGERDYQDLMEVVDEALRRFDFIDGERLGVTGGSYGGFMTNW-IVGHTNRFKAAVTQRS 515

Qy         349 VTLWIFY----DTGY 359
               :: || :    | ||
Db         516 ISNWISFFGTTDIGY 530


RESULT 14
US-09-355-166-1
; Sequence 1, Application US/09355166
; Patent No. 6316241
; GENERAL INFORMATION:
;  APPLICANT:  Genencor International, Inc.
;  TITLE OF INVENTION:  Alpha/Beta Hydrolase-Fold Enzymes
;  FILE REFERENCE:  GC511-PCT
;  CURRENT APPLICATION NUMBER: US/09/355,166
;  CURRENT FILING DATE:  1999-07-20
;  NUMBER OF SEQ ID NOS:  21
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 657
;   TYPE: PRT
;   ORGANISM: Bacillus
US-09-355-166-1

  Query Match             10.1%;  Score 196;  DB 4;  Length 657;
  Best Local Similarity   24.9%;  Pred. No. 9.1e-12;
  Matches   69;  Conservative   43;  Mismatches  107;  Indels   58;  Gaps    9;

Qy         111 TKDSPLEHHLYVVSYVNPGEVTRLTDRG---------------------YSHSCCISQHC 149
               ||||      ||:           ||:|                     | :|  :|
Db         311 TKDS---QGFYVIG----------TDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDE 357

Qy         150 DFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYT--PPEIFSFE 207
                 ||:  :   | :   |   |   |   :|          : ::|   ||   :
Db         358 QHFIASVTKPDRPSELYSIPLGQEEKQLTGANDKF----------VREHTISIPEEIQYA 407

Qy         208 STTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYF-RLNTLASLGYV 266
               :  |  : | | :|  ::    || :| |:||| :        |  ||     ||: ||
Db         408 TEDGVMVNGWLMRPAQMEGETTYPLILNIHGGPHMMY------GHTYFHEFQVLAAKGYA 461

Qy         267 VVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGG 326
               || |: |||   | :|  | :   |  : || :: :     |   ||  |:|: | ||||
Db         462 VVYIMPRGSHGYGQEFVNAVRGDYGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGG 521

Qy         327 YLSLMALMQRSDIFRVAIAGAPVTLWIFY----DTGY 359
               :::   :: ::: |: |:    :: || :    | ||
Db         522 FMTNW-IVGQTNRFKAAVTQRSISNWISFHGVSDIGY 557


RESULT 15
US-08-664-646A-2
; Sequence 2, Application US/08664646A
; Patent No. 5877001
;  GENERAL INFORMATION:
;    APPLICANT:  Murphy et al.
;    TITLE OF INVENTION:  Amidases
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/664,646A
;      FILING DATE:  June 17, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles J. Herron
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-53
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  622 AMINO ACIDS
;      TYPE:  AMINO ACID
;      STRANDEDNESS:
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  PROTEIN
US-08-664-646A-2

  Query Match             8.6%;  Score 166.5;  DB 2;  Length 622;
  Best Local Similarity   29.6%;  Pred. No. 1.1e-08;
  Matches   48;  Conservative   33;  Mismatches   68;  Indels   13;  Gaps    6;

Qy         201 PEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTL 260
               |  | |:|     : |   :|   :  :| | ::|::|||:  :  :||     : :  :
Db         352 PRHFRFKS-KDLEIDGWYLRPEVKE--EKAPVIVFVHGGPK-GMYGHRF----VYEMQLM 403

Qy         261 ASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIH 320
               || || || :: |||     |    : |  : :| : |::      | :||||
Db         404 ASKGYYVVFVNPRGSDGYSEDFALRVLERTGLEDFEDIMNGIEEFFKLEPQADRERVGIT 463

Qy         321 GWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWI----FYDTG 358
               | ||||:::  || | ||:|:  |:   :: |:    | | |
Db         464 GISYGGFMTNWALTQ-SDLFKAGISENGISYWLTSYAFSDIG 504


Search completed: October 15, 2003, 17:15:40
Job time : 10.0317 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:13:57 ; Search time 50.8676 Seconds
(without alignments)
1140.337 Million cell updates/sec

Title: US-10-070-464-7
Perfect score: 1933
Sequence: 1 EEDARSAGVATFVLQEEFDR.....RVAIAGAPVTLWIFYDTGYT 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 600653 seqs, 161128416 residues

Total number of hits satisfying chosen parameters: 600653

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1836.5	95.0	882	10	US-09-976-674-1	Sequence 1, Appli
2	1836.5	95.0	882	12	US-10-054-776-2	Sequence 2, Appli
3	1836.5	95.0	882	12	US-10-170-789-38	Sequence 38, Appl
4	1220.5	63.1	863	10	US-09-976-674-3	Sequence 3, Appli
5	1220.5	63.1	892	10	US-09-976-674-23	Sequence 23, Appl
6	1220.5	63.1	892	10	US-09-976-674-27	Sequence 27, Appl
7	1201	62.1	879	10	US-09-976-674-33	Sequence 33, Appl
8	1201	62.1	879	10	US-09-976-674-35	Sequence 35, Appl
9	1199.5	62.1	658	10	US-09-976-674-19	Sequence 19, Appl
10	1199.5	62.1	661	10	US-09-976-674-11	Sequence 11, Appl
11	1199.5	62.1	690	10	US-09-976-674-7	Sequence 7, Appli
12	1132.5	58.6	832	10	US-09-976-674-29	Sequence 29, Appl
13	1132.5	58.6	832	10	US-09-976-674-31	Sequence 31, Appl
14	1113	57.6	819	10	US-09-976-674-37	Sequence 37, Appl
15	1113	57.6	819	10	US-09-976-674-39	Sequence 39, Appl
16	1086.5	56.2	310	12	US-09-993-959-4	Sequence 4, Appli
17	944	48.8	508	15	US-10-237-271-3	Sequence 3, Appli
18	931.5	48.2	613	10	US-09-976-674-21	Sequence 21, Appl
19	505	26.1	358	10	US-09-976-674-13	Sequence 13, Appl
20	368	19.0	129	11	US-09-764-891-3564	Sequence 3564, Ap
21	352	18.2	518	10	US-09-976-674-25	Sequence 25, Appl
22	343.5	17.8	432	15	US-10-237-271-4	Sequence 4, Appli
23	315	16.3	766	9	US-09-265-606-3	Sequence 3, Appli
24	315	16.3	766	12	US-09-993-959-1	Sequence 1, Appli
25	315	16.3	766	12	US-10-423-714-6	Sequence 6, Appli
26	315	16.3	766	14	US-10-002-593-6	Sequence 6, Appli
27	315	16.3	766	15	US-10-165-603-7	Sequence 7, Appli
28	308.5	16.0	710	15	US-10-156-761-10681	Sequence 10681, A
29	296.5	15.3	767	15	US-10-165-603-4	Sequence 4, Appli
30	280.5	14.5	818	12	US-10-401-437-3	Sequence 3, Appli
31	280.5	14.5	818	12	US-10-402-312-3	Sequence 3, Appli
32	280.5	14.5	818	12	US-10-402-067-3	Sequence 3, Appli
33	280.5	14.5	818	12	US-10-401-436-3	Sequence 3, Appli
34	257	13.3	760	9	US-09-265-606-2	Sequence 2, Appli
35	257	13.3	760	12	US-10-301-822-55	Sequence 55, Appl
36	257	13.3	760	15	US-10-177-293-136	Sequence 136, App
37	212	11.0	803	12	US-10-205-219-58	Sequence 58, Appl
38	197.5	10.2	706	10	US-09-976-674-41	Sequence 41, Appl
39	197.5	10.2	796	10	US-09-976-674-5	Sequence 5, Appli
40	197.5	10.2	796	12	US-09-870-133-2	Sequence 2, Appli
41	197.5	10.2	796	15	US-10-160-501-5	Sequence 5, Appli
42	196	10.1	657	9	US-09-950-368-1	Sequence 1, Appli
43	196	10.1	657	12	US-10-401-437-2	Sequence 2, Appli
44	196	10.1	657	12	US-10-402-312-2	Sequence 2, Appli
45	196	10.1	657	12	US-10-402-067-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-976-674-1
; Sequence 1, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-1

Query Match 95.0%; Score 1836.5; DB 10; Length 882;
Best Local Similarity 67.5%; Pred. No. 5.7e-179;
Matches 360; Conservative 0; Mismatches 0; Indels 173; Gaps 1;

```
Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 303

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              ||||||||||||||||||||||||||||||||||||
Db        304 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 363

Qy         99 ---------------------------------------------------------- 98

Db        364 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 423
```

```
Qy          99 ------------------------------------------------------------ 98

Db         424 EETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 483

Qy          99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                             |||||||||||||||||||||||||||||
Db         484 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 543

Qy         128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         544 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 603

Qy         188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 663

Qy         248 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         664 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 723

Qy         308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         724 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 776


RESULT 2
US-10-054-776-2
; Sequence 2, Application US/10054776
; Publication No..US20030165818A1
; GENERAL INFORMATION:
;  APPLICANT: Mark Robert Edbrooke
;  APPLICANT:  Alan Peter Lewis
;  TITLE OF INVENTION: NOVEL PROTEIN
;  FILE REFERENCE: QG1042US
;  CURRENT APPLICATION NUMBER: US/10/054,776
;  CURRENT FILING DATE:  2002-01-23
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 882
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-054-776-2

  Query Match             95.0%;  Score 1836.5;  DB 12;  Length 882;
  Best Local Similarity   67.5%;  Pred. No. 5.7e-179;
  Matches  360;  Conservative    0;  Mismatches    0;  Indels  173;  Gaps    1;

Qy           1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         244 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 303

Qy          61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
               ||||||||||||||||||||||||||||||||||||
Db         304 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 363

Qy          99 ------------------------------------------------------------ 98

Db         364 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 423

Qy          99 ------------------------------------------------------------ 98

Db         424 EETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 483

Qy          99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                             |||||||||||||||||||||||||||||
Db         484 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 543

Qy         128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         544 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 603

Qy         188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 663

Qy         248 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         664 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 723

Qy         308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         724 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 776


RESULT 3
US-10-170-789-38
; Sequence 38, Application US/10170789
; Publication No. US20030180930A1
; GENERAL INFORMATION:
;  APPLICANT: Rachel E. Meyers
;  APPLICANT:  Olandt, Peter J.
;  APPLICANT:  Kapeller-Libermann, Rosana
;  APPLICANT:  Curtis , Rory A. J.
;  APPLICANT:  Williamson, Mark
;  APPLICANT:  Weich, Nadine
;  TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,
;  TITLE OF INVENTION:  AND PROTEASE FAMILY MEMBERS AND USES THEREOF
;  FILE REFERENCE: 10448-191001
;  CURRENT APPLICATION NUMBER: US/10/170,789
;  CURRENT FILING DATE:  2002-06-13
;  PRIOR APPLICATION NUMBER: US 09/797,039
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: PCT/US01/06525
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: US 60/186,061
;  PRIOR FILING DATE: 2000-02-29
;  PRIOR APPLICATION NUMBER: US 09/882,166
;  PRIOR FILING DATE: 2001-06-15
;  PRIOR APPLICATION NUMBER: PCT/US01/19269
;  PRIOR FILING DATE: 2001-06-15
;  PRIOR APPLICATION NUMBER: US 60/212,078
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 09/934,406
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: PCT/US01/26052
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: US 60/226,740
;  PRIOR FILING DATE: 2000-08-21
;  PRIOR APPLICATION NUMBER: US 09/861,801
;  PRIOR FILING DATE: 2001-05-21
;  PRIOR APPLICATION NUMBER: PCT/US01/16549
;  PRIOR FILING DATE: 2001-05-21
;  PRIOR APPLICATION NUMBER: US 60/205,508
;  PRIOR FILING DATE: 2000-05-19
;  PRIOR APPLICATION NUMBER: US 09/801,267
;  PRIOR FILING DATE: 2001-03-06
;  PRIOR APPLICATION NUMBER: PCT/US01/07138
;  PRIOR FILING DATE: 2001-03-05
;  PRIOR APPLICATION NUMBER: US 60/187,454
;  PRIOR FILING DATE: 2000-03-07
;  PRIOR APPLICATION NUMBER: US 09/829,671
;  PRIOR FILING DATE: 2001-04-10
;  PRIOR APPLICATION NUMBER: PCT/US01/40483
;  PRIOR FILING DATE: 2001-04-11
;  PRIOR APPLICATION NUMBER: US 60/197,508
;  PRIOR FILING DATE: 2000-04-18
;  PRIOR APPLICATION NUMBER: US 09/961,721
;  PRIOR FILING DATE: 2001-09-24
;  PRIOR APPLICATION NUMBER: PCT/US01/29904
;  PRIOR FILING DATE: 2001-09-24
;  PRIOR APPLICATION NUMBER: US 60/235,023
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US 10/045,367
;  PRIOR FILING DATE: 2001-11-07
```

```
;  PRIOR APPLICATION NUMBER: US 60/246,561
;  PRIOR FILING DATE: 2000-11-07
;  PRIOR APPLICATION NUMBER: US 09/801,275
;  PRIOR FILING DATE: 2001-03-06
;  PRIOR APPLICATION NUMBER: PCT/US01/07074
;  PRIOR FILING DATE: 2001-03-05
;  PRIOR APPLICATION NUMBER: US 60/187,420
;  PRIOR FILING DATE: 2000-03-07
;  NUMBER OF SEQ ID NOS: 63
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 38
;   LENGTH: 882
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-170-789-38

  Query Match             95.0%;  Score 1836.5;  DB 12;  Length 882;
  Best Local Similarity   67.5%;  Pred. No. 5.7e-179;
  Matches  360;  Conservative    0;  Mismatches    0;  Indels  173;  Gaps    1;

Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 303

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              ||||||||||||||||||||||||||||||||||||
Db        304 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 363

Qy         99 ---------------------------------------------------------- 98

Db        364 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 423

Qy         99 ---------------------------------------------------------- 98

Db        424 EETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 483

Qy         99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                           ||||||||||||||||||||||||||||||
Db        484 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 543

Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 603

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 663

Qy        248 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 723

Qy        308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 776


RESULT 4
US-09-976-674-3
; Sequence 3, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 863
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-976-674-3

  Query Match             63.1%;  Score 1220.5;  DB 10;  Length 863;
  Best Local Similarity   44.2%;  Pred. No. 7.7e-116;
  Matches  236;  Conservative   54;  Mismatches   69;  Indels  175;  Gaps    2;

Qy          2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
              :| :||||||||:|||||||:||||||  |    | | | ||||||| ||||||:||| ||
Db        234 DDPKSAGVATFVIQEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSP 293

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
                || |: ||:|||:||: |||:  |::|   |::|:|:
Db        294 ALEERKTDSYRYPRTGSKNPKIALKLAEFQTDSQGKIVSTQEKELVQPFSSLFPKVEYIA 353

Qy         99 ---------------------------------------------------------- 98

Db        354 RAGWTRDGKYAWAMFLDRPQQWLQLVLLPPALFIPSTENEEQRLASARAVPRNVQPYVVY 413

Qy         99 ---------------------------------------------------------- 98

Db        414 EEVTNVWINVHDIFYPFPQSEGEDELCFLRANECKTGFCHLYKVTAVLKSQGYDWSEPFS 473

Qy         99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYV 126
                                           |:|  :||||:||||:||||||||||
Db        474 PGEDEFKCPIKEEIALTSGEWEVLARHGSKIWVNEETKLVYFQGTKDTPLEHHLYVVSYE 533

Qy        127 NPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWA 186
                ||: |||  |:|||| :||: | |:| ||:   | || :|||| |:||| |   |||
Db        534 AAGEIVRLTTPGFSHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWA 593

Qy        187 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVN 246
              :::::|   ||| ||||| | : :   ||||:|||| ||||||:|||||:||||||||||
Db        594 SMMEAASCPPDYVPPEIFHFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVN 653

Qy        247 NRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLA 306
              | |||:|| |||||||||| ||||| |||| |||:|||| | :|||:||:|||||||:|
Db        654 NSFKGIKYLRLNTLASLGYAVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVA 713

Qy        307 SRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              :| |||| || ||||||||:|||| |: :  :|:||||||||:|: ||||||
Db        714 EKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVWMAYDTGYT 767


RESULT 5
US-09-976-674-23
; Sequence 23, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
;   LENGTH: 892
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-976-674-23

  Query Match             63.1%;  Score 1220.5;  DB 10;  Length 892;
  Best Local Similarity   44.2%;  Pred. No. 8.1e-116;
```

Matches 236; Conservative 54; Mismatches 69; Indels 175; Gaps 2;

```
Qy          2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
              :| :||||||||:||||||::|||||| |    | | | ||||||| ||||||:||| ||
Db        263 DDPKSAGVATFVIQEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSP 322

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------- 98
               || |: ||:|||:||: |||:  |::|   |::|:|:
Db        323 ALEERKTDSYRYPRTGSKNPKIALKLAEFQTDSQGKIVSTQEKELVQPFSSLFPKVEYIA 382

Qy         99 ------------------------------------------------------------ 98

Db        383 RAGWTRDGKYAWAMFLDRPQQWLQLVLLPPALFIPSTENEEQRLASARAVPRNVQPYVVY 442

Qy         99 ------------------------------------------------------------ 98

Db        443 EEVTNVWINVHDIFYPFPQSEGEDELCFLRANECKTGFCHLYKVTAVLKSQGYDWSEPFS 502

Qy         99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYV 126
                                            |:|  :||||:||||:||||||||||
Db        503 PGEDEFKCPIKEEIALTSGEWEVLARHGSKIWVNEETKLVYFQGTKDTPLEHHLYVVSYE 562

Qy        127 NPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWA 186
                ||: |||  |:|||| :||: | |:| ||:   | || :|||| |:|||  |   |||
Db        563 AAGEIVRLTTPGFSHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWA 622

Qy        187 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVN 246
                :::::|   ||| ||||| | : :   ||||:|||| ||||||:|||||:||||||||||
Db        623 SMMEAASCPPDYVPPEIFHFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVN 682

Qy        247 NRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLA 306
              | |||:|| |||||||||| ||||| |||| |||:|||| | :|||:||:|||||||::|
Db        683 NSFKGIKYLRLNTLASLGYAVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVA 742

Qy        307 SRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              :| |||| || ||||||||:|||| |: :  :|:|||||||||:|: ||||||
Db        743 EKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVWMAYDTGYT 796
```

RESULT 6
US-09-976-674-27
; Sequence 27, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-27

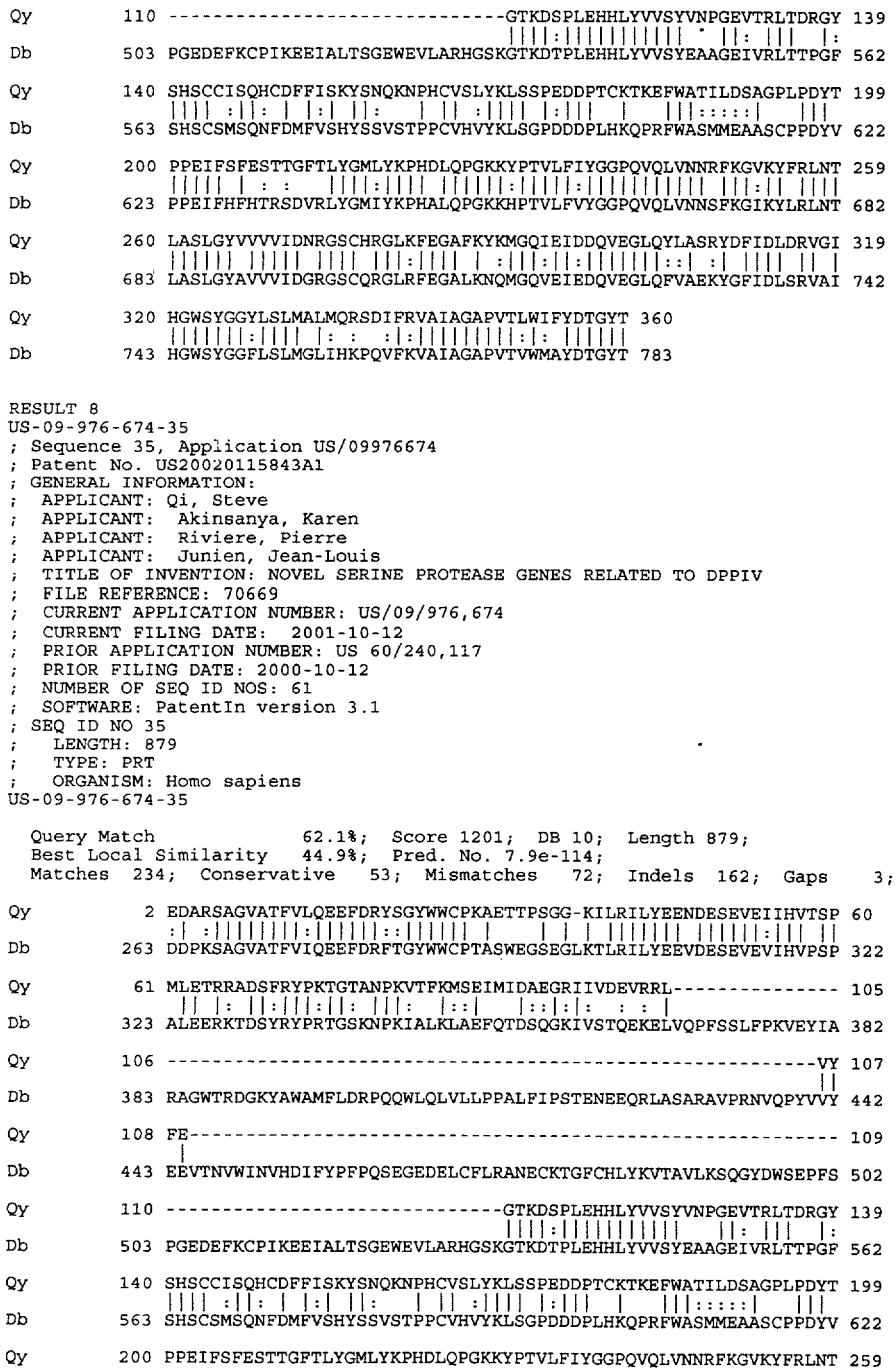
Query Match 63.1%; Score 1220.5; DB 10; Length 892;
Best Local Similarity 44.2%; Pred. No. 8.1e-116;
Matches 236; Conservative 54; Mismatches 69; Indels 175; Gaps 2;

```
Qy          2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
              :| :||||||||:||||||::|||||| |    | | | ||||||| ||||||:||| ||
Db        263 DDPKSAGVATFVIQEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSP 322

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------- 98
               || |: ||:|||:||: |||:  |::|   |::|:|:
Db        323 ALEERKTDSYRYPRTGSKNPKIALKLAEFQTDSQGKIVSTQEKELVQPFSSLFPKVEYIA 382

Qy         99 ------------------------------------------------------------ 98

Db        383 RAGWTRDGKYAWAMFLDRPQQWLQLVLLPPALFIPSTENEEQRLASARAVPRNVQPYVVY 442

Qy         99 ------------------------------------------------------------ 98

Db        443 EEVTNVWINVHDIFYPFPQSEGEDELCFLRANECKTGFCHLYKVTAVLKSQGYDWSEPFS 502

Qy         99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYV 126
                                            |:|  :||||:||||:||||||||||
Db        503 PGEDEFKCPIKEEIALTSGEWEVLARHGSKIWVNEETKLVYFQGTKDTPLEHHLYVVSYE 562

Qy        127 NPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWA 186
                ||: |||  |:|||| :||: | |:| ||:   | || :|||| |:|||  |   |||
Db        563 AAGEIVRLTTPGFSHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWA 622

Qy        187 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVN 246
                :::::|   ||| ||||| | : :   ||||:|||| ||||||:|||||:||||||||||
Db        623 SMMEAASCPPDYVPPEIFHFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVN 682

Qy        247 NRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLA 306
              | |||:|| |||||||||| ||||| |||| |||:|||| | :|||:||:|||||||::|
Db        683 NSFKGIKYLRLNTLASLGYAVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVA 742

Qy        307 SRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              :| |||| || ||||||||:|||| |: :  :|:|||||||||:|: ||||||
Db        743 EKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVWMAYDTGYT 796
```

RESULT 7
US-09-976-674-33
; Sequence 33, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-33

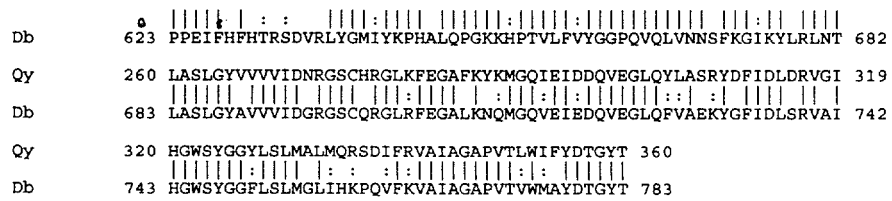
Query Match 62.1%; Score 1201; DB 10; Length 879;
Best Local Similarity 44.9%; Pred. No. 7.9e-114;
Matches 234; Conservative 53; Mismatches 72; Indels 162; Gaps 3;

```
Qy          2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
              :| :||||||||:||||||::|||||| |    | | | ||||||| ||||||:||| ||
Db        263 DDPKSAGVATFVIQEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSP 322

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIVDEVRRL---------- 105
               || |: ||:|||:||: |||:  |::|   |::|:|:  : : |
Db        323 ALEERKTDSYRYPRTGSKNPKIALKLAEFQTDSQGKIVSTQEKELVQPFSSLFPKVEYIA 382

Qy        106 ------------------------------------------------------VY 107
                                                                     ||
Db        383 RAGWTRDGKYAWAMFLDRPQQWLQLVLLPPALFIPSTENEEQRLASARAVPRNVQPYVVY 442

Qy        108 FE---------------------------------------------------------- 109
              |
Db        443 EEVTNVWINVHDIFYPFPQSEGEDELCFLRANECKTGFCHLYKVTAVLKSQGYDWSEPFS 502
```



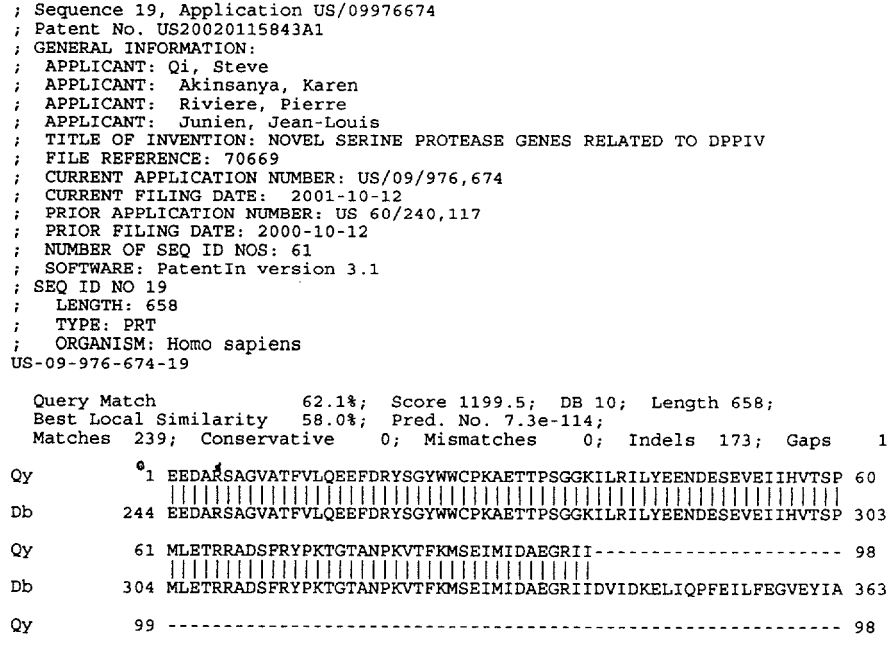
RESULT 8
US-09-976-674-35
; Sequence 35, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-35

Query Match 62.1%; Score 1201; DB 10; Length 879;
Best Local Similarity 44.9%; Pred. No. 7.9e-114;
Matches 234; Conservative 53; Mismatches 72; Indels 162; Gaps 3;



RESULT 9
US-09-976-674-19
; Sequence 19, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-19

Query Match 62.1%; Score 1199.5; DB 10; Length 658;
Best Local Similarity 58.0%; Pred. No. 7.3e-114;
Matches 239; Conservative 0; Mismatches 0; Indels 173; Gaps 1;



RESULT 10
US-09-976-674-11
; Sequence 11, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:

```
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
;   LENGTH: 661
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-976-674-11

  Query Match             62.1%;  Score 1199.5;  DB 10;  Length 661;
  Best Local Similarity   58.0%;  Pred. No. 7.4e-114;
  Matches  239;  Conservative    0;  Mismatches    0;  Indels  173;  Gaps    1;

Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 303

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              ||||||||||||||||||||||||||||||||||||
Db        304 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 363

Qy         99 ---------------------------------------------------------- 98

Db        364 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 423

Qy         99 ---------------------------------------------------------- 98

Db        424 EETTDIWININHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 483

Qy         99 -------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                             ||||||||||||||||||||||||||||
Db        484 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 543

Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 603

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 239
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 655


RESULT 11
US-09-976-674-7
; Sequence 7, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;   LENGTH: 690
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-976-674-7

  Query Match             62.1%;  Score 1199.5;  DB 10;  Length 690;
  Best Local Similarity   58.0%;  Pred. No. 7.9e-114;
  Matches  239;  Conservative    0;  Mismatches    0;  Indels  173;  Gaps    1;

Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 303

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              ||||||||||||||||||||||||||||||||||||
Db        304 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 363

Qy         99 ---------------------------------------------------------- 98

Db        364 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 423

Qy         99 ---------------------------------------------------------- 98

Db        424 EETTDIWININHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 483

Qy         99 -------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                             ||||||||||||||||||||||||||||
Db        484 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 543

Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 603

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 239
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 655


RESULT 12
US-09-976-674-29
; Sequence 29, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
;   LENGTH: 832
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-976-674-29

  Query Match             58.6%;  Score 1132.5;  DB 10;  Length 832;
  Best Local Similarity   42.3%;  Pred. No. 7.7e-107;
  Matches  222;  Conservative   52;  Mismatches   76;  Indels  175;  Gaps    2;

Qy          2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
              :| :|||||||:||||||::|||||| |    | | | ||||||| ||||||:||| ||
Db        263 DDPKSAGVATFVIQEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSP 322

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
               || |: ||:|||:||: |||:  |::|   |::|:|:
Db        323 ALEERKTDSYRYPRTGSKNPKIALKLAEFQTDSQGKIVSTQEKELVQPFSSLFPKVEYIA 382

Qy         99 ---------------------------------------------------------- 98

Db        383 RAGWTRDGKYAWAMFLDRPQQWLQLVLLPPALFIPSTENEEQRLASARAVPRNVQPYVVY 442
```

```
Qy          99 ------------------------------------------------------------ 98

Db         443 EEVTNVWINVHDIFYPFPQSEGEDELCFLRANECKTGFCHLYKVTAVLKSQGYDWSEPFS 502

Qy          99 --------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYV 126
                                               |:|  :||||:||||:||||||||||
Db         503 PGEDEFKCPIKEEIALTSGEWEVLARHGSKIWVNEETKLVYFQGTKDTPLEHHLYVVSYE 562

Qy         127 NPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWA 186
                 ||: |||  |:|||| :||: | |:| ||:   | || :|||| |:|||  |   |||
Db         563 AAGEIVRLTTPGFSHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWA 622

Qy         187 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVN 246
                 :::::|   ||| ||||| | : :   ||||:|||| ||||||:|||||:||||||||||
Db         623 SMMEAASCPPDYVPPEIFHFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVN 682

Qy         247 NRFKGVKYFRLNTLASLGYVVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLA 306
                 | |||:|| |||||||||| ||||| |||| |||:|||| | :|||:||:|||||||::|
Db         683 NSFKGIKYLRLNTLASLGYAVVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVA 742

Qy         307 SRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTL 351
                  :| |||| || ||||||||:|||| |: :  :|:       |  |
Db         743 EKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKAQPLAYPPRL 787


RESULT 13
US-09-976-674-31
; Sequence 31, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
;   LENGTH: 832
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-976-674-31

  Query Match             58.6%;  Score 1132.5;  DB 10;  Length 832;
  Best Local Similarity   42.3%;  Pred. No. 7.7e-107;
  Matches  222;  Conservative   52;  Mismatches   76;  Indels  175;  Gaps    2;

Qy           2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
                 :| :||||||||:||||||::|||||| |    | | | ||||||| ||||||:||| ||
Db         263 DDPKSAGVATFVIQEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSP 322

Qy          61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
                 || |: ||:|||:||: |||:  |::|   |::|:|:
Db         323 ALEERKTDSYRYPRTGSKNPKIALKLAEFQTDSQGKIVSTQEKELVQPFSSLFPKVEYIA 382

Qy          99 ----------------------------------------------------------- 98

Db         383 RAGWTRDGKYAWAMFLDRPQQWLQLVLLPPALFIPSTENEEQRLASARAVPRNVQPYVVY 442

Qy          99 ------------------------------------------------------------ 98

Db         443 EEVTNVWINVHDIFYPFPQSEGEDELCFLRANECKTGFCHLYKVTAVLKSQGYDWSEPFS 502

Qy          99 --------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYV 126
                                               |:|  :||||:||||:||||||||||
Db         503 PGEDEFKCPIKEEIALTSGEWEVLARHGSKIWVNEETKLVYFQGTKDTPLEHHLYVVSYE 562

Qy         127 NPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWA 186
                 ||: |||  |:|||| :||: | |:| ||:   | || :|||| |:|||  |   |||
Db         563 AAGEIVRLTTPGFSHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWA 622

Qy         187 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVN 246
                 :::::|   ||| ||||| | : :   ||||:|||| ||||||:|||||:||||||||||
Db         623 SMMEAASCPPDYVPPEIFHFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVN 682

Qy         247 NRFKGVKYFRLNTLASLGYVVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLA 306
                 | |||:|| |||||||||| ||||| |||| |||:|||| | :|||:||:|||||||::|
Db         683 NSFKGIKYLRLNTLASLGYAVVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVA 742

Qy         307 SRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTL 351
                  :| |||| || ||||||||:|||| |: :  :|:       |  |
Db         743 EKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKAQPLAYPPRL 787


RESULT 14
US-09-976-674-37
; Sequence 37, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37
;   LENGTH: 819
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-976-674-37

  Query Match             57.6%;  Score 1113;  DB 10;  Length 819;
  Best Local Similarity   43.0%;  Pred. No. 7.5e-105;
  Matches  220;  Conservative   51;  Mismatches   79;  Indels  162;  Gaps    3;

Qy           2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
                 :| :||||||||:||||||::|||||| |    | | | ||||||| ||||||:||| ||
Db         263 DDPKSAGVATFVIQEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSP 322

Qy          61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIVDEVRRL---------------- 105
                 || |: ||:|||:||: |||:  |::|   |::|:|:  : : |
Db         323 ALEERKTDSYRYPRTGSKNPKIALKLAEFQTDSQGKIVSTQEKELVQPFSSLFPKVEYIA 382

Qy         106 ----------------------------------------------------------VY 107
                                                                           ||
Db         383 RAGWTRDGKYAWAMFLDRPQQWLQLVLLPPALFIPSTENEEQRLASARAVPRNVQPYVVY 442

Qy         108 FE------------------------------------------------------------ 109
                 |
Db         443 EEVTNVWINVHDIFYPFPQSEGEDELCFLRANECKTGFCHLYKVTAVLKSQGYDWSEPFS 502

Qy         110 ----------------------------GTKDSPLEHHLYVVSYVNPGEVTRLTDRGY 139
                                            ||||:|||||||||||   ||: |||  |:
Db         503 PGEDEFKCPIKEEIALTSGEWEVLARHGSKGTKDTPLEHHLYVVSYEAAGEIVRLTTPGF 562

Qy         140 SHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYT 199
                 |||| :||: | |:| ||:   | || :|||| |:|||  |   |||:::::|   |||
Db         563 SHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWASMMEAASCPPDYV 622

Qy         200 PPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNT 259
                 ||||| | : :   ||||:|||| ||||||:|||||:||||||||||| |||:|| ||||
Db         623 PPEIFHFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNT 682

Qy         260 LASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGI 319
                 |||||| ||||| |||| |||:|||| | :|||:||:|||||||::| :| |||| || |
```

```
Db        683 LASLGYAVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAI 742

Qy        320 HGWSYGGYLSLMALMQRSDIFRVAIAGAPVTL 351
              |||||||:|||| |: :  :|:      |  |
Db        743 HGWSYGGFLSLMGLIHKPQVFKAQPLAYPPRL 774


RESULT 15
US-09-976-674-39
; Sequence 39, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
;   LENGTH: 819
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-976-674-39

  Query Match             57.6%;  Score 1113;  DB 10;  Length 819;
  Best Local Similarity   43.0%;  Pred. No. 7.5e-105;
  Matches  220;  Conservative   51;  Mismatches   79;  Indels  162;  Gaps    3;

Qy          2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
              :| :||||||||:||||||::|||||| |    | | | ||||||| ||||||:||| ||
Db        263 DDPKSAGVATFVIQEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSP 322

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIVDEVRRL---------------- 105
               || |: ||:|||:||: |||:  |::|   |::|:|:  : : |
Db        323 ALEERKTDSYRYPRTGSKNPKIALKLAEFQTDSQGKIVSTQEKELVQPFSSLFPKVEYIA 382

Qy        106 ------------------------------------------------------------VY 107
                                                                           ||
Db        383 RAGWTRDGKYAWAMFLDRPQQWLQLVLLPPALFIPSTENEEQRLASARAVPRNVQPYVVY 442

Qy        108 FE---------------------------------------------------------- 109
               |
Db        443 EEVTNVWINVHDIFYPFPQSEGEDELCFLRANECKTGFCHLYKVTAVLKSQGYDWSEPFS 502

Qy        110 -------------------------------GTKDSPLEHHLYVVSYVNPGEVTRLTDRGY 139
                                            ||||:|||||||||   ||: |||  |:
Db        503 PGEDEFKCPIKEEIALTSGEWEVLARHGSKGTKDTPLEHHLYVVSYEAAGEIVRLTTPGF 562

Qy        140 SHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYT 199
              |||| :||: | |:| ||:   | || :|||| |:||| |   |||:::::|   |||
Db        563 SHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWASMMEAASCPPDYV 622

Qy        200 PPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNT 259
              ||||| | : :   ||||:|||| |||||:|||||:|||||||||| |||:|| ||||
Db        623 PPEIFHFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNT 682

Qy        260 LASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGI 319
              |||||| ||||| |||| |||:|||| | :|||:||:|||||||::| :| |||| || |
Db        683 LASLGYAVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAI 742

Qy        320 HGWSYGGYLSLMALMQRSDIFRVAIAGAPVTL 351
              |||||||:|||| |: :  :|:      |  |
Db        743 HGWSYGGFLSLMGLIHKPQVFKAQPLAYPPRL 774


Search completed: October 15, 2003, 17:28:53
Job time : 52.8676 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:08:55 ; Search time 9.45959 Seconds
(without alignments)
3659.853 Million cell updates/sec

Title: US-10-070-464-7
Perfect score: 1933
Sequence: 1 EEDARSAGVATFVLQEEFDR.....RVAIAGAPVTLWIFYDTGYT 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	527.5	27.3	738	2	A87516	dipeptidyl peptida
2	491.5	25.4	931	2	T32919	hypothetical prote
3	430.5	22.3	741	2	JC5142	X-Pro dipeptidyl-p
4	369	19.1	711	2	S66261	X-Pro dipeptidyl-p
5	315	16.3	766	1	CDHU26	dipeptidyl-peptida
6	312.5	16.2	793	2	T41703	dipeptidyl aminope
7	309	16.0	760	1	S23752	dipeptidyl-peptida
8	291.5	15.1	792	1	A39914	dipeptidyl-peptida
9	280.5	14.5	818	1	A30107	dipeptidyl aminope
10	250	12.9	931	2	A49737	dipeptidyl aminope
11	245	12.7	795	2	F82858	dipeptidyl-peptida
12	240	12.4	829	2	T19514	hypothetical prote
13	226.5	11.7	683	2	E87495	prolyl oligopeptid
14	221	11.4	803	2	A41793	dipeptidyl aminope
15	220	11.4	709	2	B82580	alanyl dipeptidyl
16	219	11.3	779	2	T25173	hypothetical prote
17	219	11.3	799	2	T25174	hypothetical prote
18	205.5	10.6	631	2	H75007	probable acylamino
19	204	10.6	642	2	C71137	hypothetical prote
20	203	10.5	803	2	I68600	dipeptidyl aminope
21	203	10.5	865	2	I54331	dipeptidyl aminope
22	199	10.3	622	2	F71174	hypothetical prote
23	197.5	10.2	743	2	T37700	probable dipeptidy
24	196	10.1	657	2	E70025	probable acylamino
25	193	10.0	632	2	E75057	peptidase PAB1418
26	189	9.8	536	2	F90299	acylaminoacyl-pept
27	174	9.0	759	2	I38593	fibroblast activat
28	166.5	8.6	674	2	B84381	acylaminoacyl-pept
29	165	8.5	591	2	H72474	probable acylamino
30	162.5	8.4	569	2	S74053	probable acylamino
31	149	7.7	676	2	C97775	acylamino-acid-rel
32	147.5	7.6	667	2	A87711	prolyl oligopeptid
33	147.5	7.6	721	2	T09631	probable acylamino
34	140.5	7.3	591	2	A99444	acylaminoacyl-pept
35	140.5	7.3	606	2	T35378	probable peptide h
36	139.5	7.2	745	2	T33751	hypothetical prote
37	138.5	7.2	598	2	P84199	hypothetical prote
38	137.5	7.1	654	2	AD3183	peptidase [importe
39	134.5	7.0	732	1	JC4655	acylaminoacyl-pept
40	131.5	6.8	732	1	S07624	acylaminoacyl-pept
41	130.5	6.8	572	2	F72455	probable acylamino
42	129.5	6.7	555	2	AD1794	acylase and dieste
43	127.5	6.6	555	2	AB1419	acylase and dieste
44	126.5	6.5	629	2	T15945	hypothetical prote
45	126.5	6.5	689	1	JN0585	prolyl oligopeptid

ALIGNMENTS

RESULT 1
A87516
dipeptidyl peptidase IV [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: A87516
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kn, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: A87516
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-738 <STO>
A;Cross-references: GB:AE005673; NID:g13423647; PIDN:AAK24125.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC2154

Query Match 27.3%; Score 527.5; DB 2; Length 738;
Best Local Similarity 28.5%; Pred. No. 1e-36;
Matches 137; Conservative 61; Mismatches 132; Indels 151; Gaps 13;

```
Qy          2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPM 61
              :|| | ||| |::||| ||::|||| |                ||:|   ||| |:|:
Db        189 KDALSFGVAEFIVQEELDRFTGYWWSPDES---------RIVYTRVDESGVDIV------ 233

Qy         62 LETRRAD---------SFRYPKTGTANP-----------KVTF---------------- 84
                 |||          : |||: |  |                |||
Db        234 ---PRADIGPGGATVVNQRYPRAGRPNAVVDLFVRDLASGKVTALDLGANKDIYVARVAW 290

Qy         85 ----------KMSE-------------------IMID---------------------- 92
                     ::|                     |: |
Db        291 SADGKTVYVQRLSRDQKTLDLLAFDAATGAGKTILTDTDPHFIEVSNDFRPLTDGTFLWG 350

Qy         93 -------------AEGRII---------------VDEVRRLVYFEGTKDSPLEHHLYVVS 124
                          |:|::|                ||| |::  |  : |:|:|  || ||
Db        351 SEKDGNQHLYRYAADGKLIAQITKGDWPVIGLEGVDEARKVAIFSASIDTPIERRLYEVS 410

Qy        125 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSP-----EDDPTC 179
              |  ||:   ||  |   :  :: :   |   ||: | |   :||          |::
Db        411 YAKPGKPKALTSAGGWWAAKVADNGGAFAGTYSDPKTPSQTALYSADGKRVRWIEENKLA 470

Qy        180 KTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 239
              :   :|       |  ||    ||  | ::  | ||:  : ||    | |||| :: :|||
Db        471 EGHPYW----PYAANLPQ---PEFGSLKAADGETLHYEILKPIGFDPAKKYPAIVSVYGG 523

Qy        240 PQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQV 299
              |  | |   :          |    |||:  :|||||:|  ||  |   |:| :|::||:
Db        524 PHAQRVMKNWHSPSE---RTYLEAGYVIFKLDNRGSGNRSAKFMRALDRKLGTVEVEDQL 580
```

```
Qy        300 EGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGY 359
              | ::|||:  ::| |::|: ||||||:::|| |   :  |:   |||| | |  ||| |
Db        581 LGAKFLASQ-PYVDADKLGVMGWSYGGFMALMLLTAENTPFKAGAAGAPPTEWSLYDTAY 639

Qy        360 T 360
              |
Db        640 T 640
```

RESULT 2
T32919
hypothetical protein K02F2.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T32919
R;Maggi, L.; Goela, D.
submitted to the EMBL Data Library, January 1998
A;Description: The sequence of C. elegans cosmid K02F2.
A;Reference number: Z21246
A;Accession: T32919
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-931 <MAG>
A;Cross-references: EMBL:AF043699; PIDN:AAB97564.1; GSPDB:GN00019; CESP:K02F2.1
A;Experimental source: strain Bristol N2; clone K02F2
C;Genetics:
A;Gene: CESP:K02F2.1
A;Map position: 1
A;Introns: 58/3; 82/2; 131/2; 178/2; 275/3; 322/2; 404/2; 441/1; 464/2; 486/3; 528/2; 57

  Query Match 25.4%; Score 491.5; DB 2; Length 931;
  Best Local Similarity 25.9%; Pred. No. 1.6e-33;
  Matches 145; Conservative 60; Mismatches 112; Indels 243; Gaps 19;

```
Qy          8 GVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV-------TSP 60
              || ::::||| :|: | ||  :::|        |:|||  :| :|             :|
Db        298 GVPSYIVQEELERFEGIWW-SESKT--------RLLYEHVNEEKVAESQFGVNGDPPVAP 348

Qy         61 MLETRRADSFRYPKTGTANPKVTFKM---------------------------------- 86
              |         :||: || |   | :|
Db        349 M---------KYPRAGTKNAYSTLRMVILENGKAYDVPLKDEVIYKHCPFYEYITRAGFF 399

Qy         87 ------------------------------------------------------------ 86

Db        400 SDGTTVWVQVMSRDQAQCSLLLIPYTDFLLPEELGGSIKEDNLQLSTDLNMGVWDDKSHE 459

Qy         87 ------------------------------------------------------------ 86

Db        460 ETMEKPPRGKLRGTVQIHKARNDYWINTHNAIYPLKITDEEHPMYEFIYCLEKPNGSCLA 519

Qy         87 ---------------SEIMIDAEG-------RIIVDEVRRLVYFEGTKDSPLEHHLYVVS 124
                           | :: ||          |:||||| |||:   :  | | ::  ||
Db        520 LISAELDQNGYCRHTEEKLLMAENFSINKSMGIVVDEVRELVYYVANESHPTEWNI-CVS 578

Qy        125 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQK----NPHCVSLYKLS--SPEDDPT 178
              :   |:  :||: |         |  | |: :| |       | : | |  || :
Db        579 HYRTGQHAQLTESGI---------C--FKSERANGKLALDLDHGFACYMTSVGSPAE--- 624

Qy        179 CKTKEF-W------------ATILDSAGP-LPD--YTPPEIFSFES-TTGFTLYGMLYKP 221
              |:   | |            | |  |  |  ||  :  ||:  |:|  ||    | |: :|
Db        625 CRFYSFRWKENEVLPSTVYAANITVSGHPGQPDLHFDSPEMIEFQSKKTGLMHYAMILRP 684

Qy        222 HDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLK 281
              :  | ||||   ::|||| :|:|:| |  ::| |       |||||| |||||| |||::
Db        685 SNFDPYKKYPVFHYVYGGPGIQIVHNDFSWIQYIR---FCRLGYVVVFIDNRGSAHRGIE 741

Qy        282 FEGAFKYKMGQIEIDDQVEGLQYLASRY-DFIDLDRVGIHGWSYGGYLSLMALMQRSDIF 340
              ||      ||| :|::|||||||| || |   |:|: || :|||||||::|  : :  :|:
Db        742 FERHIHKKMGTVEVEDQVEGLQMLAERTGGFMDMSRVVVHGWSYGGYMALQMIAKHPNIY 801

Qy        341 RVAIAGAPVTLWIFYDTGYT 360
              | ||||  |: |  ||| ||
Db        802 RAAIAGGAVSDWRLYDTAYT 821
```

RESULT 3
JC5142
X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Xanthomonas maltophilia
C;Species: Xanthomonas maltophilia
C;Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 20-Jun-2000
C;Accession: JC5142
R;Kabashima, T.; Ito, K.; Yoshimoto, T.
J. Biochem. 120, 1111-1117, 1996
A;Title: Dipeptidyl peptidase IV from Xanthomonas maltophilia: Sequencing and express
A;Reference number: JC5142; MUID:97164011; PMID:9010758
A;Accession: JC5142
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-741 <KAB>
A;Cross-references: DDBJ:D83263; NID:g1753196; PIDN:BAA11872.1; PID:g1753197
C;Comment: This enzyme catalyzes the hydrolytic removal of amino-terminal dipeptidyl
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptide hydrolase; membrane bound
F;4-18/Domain: transmembrane #status predicted <TMM>
F;610/Active site: Ser #status predicted
F;685/Active site: Asp #status predicted
F;717/Active site: His #status predicted

  Query Match 22.3%; Score 430.5; DB 2; Length 741;
  Best Local Similarity 26.2%; Pred. No. 1.9e-28;
  Matches 125; Conservative 57; Mismatches 144; Indels 151; Gaps 13;

```
Qy          8 GVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEV------EIIHVTSPM 61
              |:| ||  || ||::|||| |           | |   ||| |      |:    : :
Db        198 GIAEFVADEEMDRHTGYWWAPDDSA---------IAYARIDESPVPVQKRYEVYADRTDV 248

Qy         62 LETRRADSFRYPKTGTANPKVTF----------------KMSEI--------------- 89
              :|       |||  | || :|                  |  :|
Db        249 IEQ------RYPAAGDANVQVKLGVISPAEQAQTQWIDLGKEQDIYLARVNWRDPQHLSF 302

Qy         90 ------------------------------------------------------------ 89

Db        303 QRQSRDQKKLDLVEVTLASNQQRVLAHETSPTWVPLHNSLRFLDDGSILWSSERTGFQHL 362

Qy         90 -MIDAEGR--------------IIVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRL 134
               ||::|:               : |||   | ||    :|  |  :| |  :  |:  ||
Db        363 YRIDSKGKAAALTHGNWSVDELLAVDEKAGLAYFRAGIESARESQIYAVP-LQGGQPQRL 421

Qy        135 TDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATIL--DSA 192
              :     ||   :::   ::  :||   |  : |:: :           |  ||::  | |
Db        422 SKAPGMHSASFARNASVYVDSWSNNSTPPQIELFRANG----------EKIATLVENDLA 471

Qy        193 GPLPDYT-------PPEIFSFESTTGFT-LYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 244
              |   |         | |  :  :  | | |   : ||    | |:||  :::|||  |
Db        472 DPKHPYARYREAQRPVEFGTLTAADGKTPLNYSVIKPAGFDPAKRYPVAVYVYGGPASQT 531

Qy        245 VNNRFKGVKYFRLNT-LASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQ 303
              | : |       |  ||  ||||  :||||:  ||  | ||   | | :|: ||: |:
Db        532 VTDSWPGRGDHLFNQYLAQQGYVVFSLDNRGTPRRGRDFGGALYGKQGTVEVADQLRGVA 591

Qy        304 YLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              :|  :  ::|  |:|: ||| |||::|| | : || :   :|||||| |  ||: ||
Db        592 WL-KQQPWVDPARIGVQGWSNGGYMTLMLLAKASDSYACGVAGAPVTDWGLYDSHYT 647
```

RESULT 4
S66261
X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Flavobacterium meningosepticum
C;Species: Flavobacterium meningosepticum
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C;Accession: S66261
R;Kabashima, T.; Yoshida, T.; Ito, K.; Yoshimoto, T.
Arch. Biochem. Biophys. 320, 123-128, 1995
A;Title: Cloning, sequencing, and expression of the dipeptidyl peptidase IV gene from
A;Reference number: S66261; MUID:95314307; PMID:7793970
A;Accession: S66261

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-711 <KAB>
A;Cross-references: EMBL:D42121; NID:g577283; PIDN:BAA07702.1; PID:g577284
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptide hydrolase

Query Match 19.1%; Score 369; DB 2; Length 711;
Best Local Similarity 23.8%; Pred. No. 3.1e-23;
Matches 112; Conservative 70; Mismatches 147; Indels 142; Gaps 13;

```
Qy          8 GVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEI-IHVTS--PMLET 64
              |:  :| :|||     | |         :|  :: : ::|    |: | |:  :  | | |
Db        172 GLGDWVYEEEFGHADYYQW------NKAGDALVFVRFDERKVPEINIPIYYQNLYPKLMT 225

Qy         65 RRADSFRYPKTGTANPKVT---------------FKMSE-------IMIDAEGRIIVDEV 102
                ::||| |  |  ||                |  ||          :|   |:|
Db        226 -----YKYPKAGEENSAVTAYLYQLSSGKSAQLNFGSSEKYYIPQLFQTNANDEIVVATA 280

Qy        103 RR-----------------LVYFEGTKDSPLE----------------------HHLY- 121
              |                    |  | :: :|                        |||
Db        281 NRHQNKVDLLKVNTKTAAVSKLFTETDNAWIETDNLTMEFLDDNSFLWASERDGHRHLYW 340

Qy        122 -----------------VVSY---------------------------VNPGEVTRLTD 136
                             :::|                           :| |:   |::
Db        341 YDAAGKLKKQVSKGDWEIINYYGYNPKTKEVYIQTTEKGSINKVVSKLNINTGKTQLLSN 400

Qy        137 RGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPE-------DDPTCKTKEFWATIL 189
               ::|   |:  ::||:  |  | |    |   :  :        ||·  | |
Db        401 AEGNNSAAFSKTFNYFINTSSTAKVPTKYILKDANGKDVKELQNNDDLLNKLKS------ 454

Qy        190 DSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRF 249
                    ::   |  :  :  |  :   : || :  | ||||  :| | ||  | | | :
Db        455 ------DNFIAKEFITIPNAAGDQMNAWMIKPKNFDPAKKYPVFMFQYSGPGSQQVANSW 508

Qy        250 KGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRY 309
              |       : ||   ||:|| :| ||:  || |::      :|: ||:||:   ::| ::
Db        509 DGGNGIWFDMLAQKGYLVVCVDGRGTGFRGTKYKKVTYKNLGKYEIEDQITAAKWLGNQ- 567

Qy        310 DFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
               ::|  |:|| |||||||:: :|: : :|:|:: || |||| | |||: ||
Db        568 SYVDKSRIGIFGWSYGGYMASLAMTKGADVFKMGIAVAPVTNWRFYDSIYT 618
```

RESULT 5
CDHU26
dipeptidyl-peptidase IV (EC 3.4.14.5) - human
N;Alternate names: cell surface glycoprotein CD26; thymocyte-activating molecule (THAM)
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 23-Aug-1996 #text_change 18-Jun-1999
C;Accession: S24313; B42408; A42408; B61136; S59510; I56154; S59857; S15520
R;Misumi, Y.; Hayashi, Y.; Arakawa, F.; Ikehara, Y.
Biochim. Biophys. Acta 1131, 333-336, 1992
A;Title: Molecular cloning and sequence analysis of human dipeptidyl peptidase IV, a ser
A;Reference number: S24313; MUID:92329551; PMID:1352704
A;Accession: S24313
A;Molecule type: mRNA
A;Residues: 1-6,'I',8-766 <MIS>
A;Cross-references: EMBL:X60708; NID:g35335; PIDN:CAA43118.1; PID:g35336
R;Darmoul, D.; Lacasa, M.; Baricault, L.; Marguet, D.; Sapin, C.; Trotot, P.; Barbat, A.
J. Biol. Chem. 267, 4824-4833, 1992
A;Title: Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon cancer
IV mRNA levels during cell differentiation.
A;Reference number: A42408; MUID:92165847; PMID:1347043
A;Accession: B42408
A;Molecule type: mRNA
A;Residues: 1-5,'R',7-436,'S',438-556,'I',558-662,'E',664-766.<DAR1>
A;Cross-references: GB:M80536; NID:g181569; PIDN:AAA52308.1; PID:g181570
A;Experimental source: intestine
A;Note: this sequence corresponds with the author's translation
A;Accession: A42408
A;Molecule type: mRNA
A;Residues: 1-5,'R',7-436,'S',438-556,'I',558-662,'E',664-711,'G',713-766 <DAR2>
A;Cross-references: GB:M80536; NID:g181569
A;Note: sequence extracted from NCBI backbone (NCBIN:83986, NCBIP:83988); this sequer
R;Gorvel, J.P.; Ferrero, A.; Chambraud, L.; Rigal, A.; Bonicel, J.; Maroux, S.
Gastroenterology 101, 618-625, 1991
A;Title: Expression of sucrase-isomaltase and dipeptidylpeptidase IV in human small
A;Reference number: A61136; MUID:91317403; PMID:1677636
A;Accession: B61136
A;Molecule type: protein
A;Residues: 1-15,'X',17-22 <GOR>
R;Boehm, S.K.; Gum Jr., J.R.; Erickson, R.H.; Hicks, J.W.; Kim, Y.S.
Biochem. J. 311, 835-843, 1995
A;Title: Human dipeptidyl peptidase IV gene promoter: tissue-specific regulation fro
A;Reference number: S59510; MUID:96067599; PMID:7487939
A;Accession: S59510
A;Molecule type: DNA
A;Residues: 1-31 <BOE>
A;Cross-references: GB:S79876; NID:g1195574; PIDN:AAB35614.1; PID:g1195575
R;Tanaka, T.; Camerini, D.; Seed, B.; Torimoto, Y.; Dang, N.H.; Kameoka, J.; Dahlberg
J. Immunol. 149, 481-486, 1992
A;Title: Cloning and functional expression of the T cell activation antigen CD26.
A;Reference number: I56154; MUID:92325476; PMID:1352530
A;Accession: I56154
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-436,'S',438-766 <TAN>
A;Cross-references: GB:M74777; NID:g180082; PIDN:AAA51943.1; PID:g180083
R;Abbott, C.A.; Baker, E.; Sutherland, G.R.; McCaughan, G.W.
Immunogenetics 40, 331-338, 1994
A;Title: Genomic organization, exact localization, and tissue expression of the humar
A;Reference number: S59857; MUID:95012454; PMID:7927537
A;Accession: S59857
A;Molecule type: DNA
A;Residues: 1-436,'S',438-766 <ABB>
A;Cross-references: EMBL:U13734
C;Genetics:
A;Gene: GDB:DPP4
A;Cross-references: GDB:125239; OMIM:102720
A;Map position: 2q24.3-2q24.3
A;Introns: 2/3; 32/1; 65/1; 95/3; 122/3; 140/2; 164/3; 205/1; 258/3; 296/2; 341/3; 35
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptide hydrolase; glycoprotein; homodimer; proteinase; transme
F;1-6/Domain: intracellular #status predicted <INT>
F;7-28/Domain: transmembrane #status predicted <TMN>
F;29-766/Domain: extracellular #status predicted <EXT>
F;85,92,150,219,229,281,321,520,685/Binding site: carbohydrate (Asn) (covalent) #stat
F;630,708,740/Active site: Ser, Asp, His #status predicted

Query Match 16.3%; Score 315; DB 1; Length 766;
Best Local Similarity 23.4%; Pred. No. 1.4e-18;
Matches 120; Conservative 59; Mismatches 146; Indels 188; Gaps 22;

```
Qy          1 EEDARSAGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH--- 56
              :||     |:  :| :|| |  ||  || |              : | : :::|| :|
Db        190 KEDIIYNGITDWVYEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSF 240

Qy         57 VTSPMLETRRADSFRYPKTGTANPKVTF-----------KMSEIMIDAEGRIIVDEVRR 104
               :   |:  :     ||| |  || | |              : | | |   :::
Db        241 YSDESLQYPKTVRVPYPKAGAVNPTVKFFVVNTDSLSSVTNATSIQITAPASMLIG---- 296

Qy        105 LVYFEGTKDSPLEHHLYVVSYVNPGEVT-----RLTDRGYSHSC----------CI--SQ 147
                        :|:|  |::     ::     |: :      |           |:   |
Db        297 ------------DHYLCDVTWATQERISLQWLRRIQNYSVMDICDYDESSGRWNCLVARQ 344

Qy        148 HCDF----FISKYSNQKNPHCV----SLYKLSSPED-------------DPTCKTKEFWA 186
              | :     :: ::     ||       | ||: | |:             | |  ||  |
Db        345 HIEMSTTGWVGRF-RPSEPHFTLDGNSFYKIISNEEGYRHICYFQIDKKDCTFITKGTWE 403

Qy        187 TI---------------------------------------------------------- 188
              |
Db        404 VIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLIDYTKVTCLSCELNPERCQYYSVSFSK 463

Qy        189 ------LDSAGP-LPDYT--------------------------PPEIFSF--ESTTG 211
                    |  :|| || ||                          | :   |   : |
```

```
Db        464 EAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSALDKMLQNVQMPSKKLDFIILNETK 523

Qy        212 FTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVV 268

Db        524 F-WYQMILPPH-FDKSKKYPLLLDVYAGPCSQ------KADTVFRLNWATYLASTENIIV 575

Qy        269 V-IDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGY 327

Db        576 ASFDGRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQF-SKMGFVDNKRIAIWGWSYGGY 634

Qy        328 LSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360

Db        635 VTSMVLGSGSGVFKCGIAVAPVSRWEYYDSVYT 667


RESULT 6
T41703
dipeptidyl aminopeptidase - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C;Accession: T41703
R;Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1998
A;Reference number: Z22011
A;Accession: T41703
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-793 <MUR>
A;Cross-references: EMBL:AL031180; PIDN:CAA20138.1; GSPDB:GN00066
A;Experimental source: strain 972h-; cosmid c2E11 -chimeric
C;Genetics:
A;Gene: SPAC2E11.08
A;Map position: 1
C;Superfamily: dipeptidyl-peptidase IV

  Query Match             16.2%;  Score 312.5;  DB 2;  Length 793;
  Best Local Similarity   22.9%;  Pred. No. 2.3e-18;
  Matches  114;  Conservative   62;  Mismatches  149;  Indels  173;  Gaps   17;

Qy          8 GVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV--------- 57

Db        214 GLTDWIYEEEVLSSPSTIWWSPDSD---------KIAFLKLNESEIPTYHYPLYTAELDP 264

Qy         58 TSPMLETRRADSFRYPKTGTANPKVTFKMSEI---------------------------- 89

Db        265 SLPEFDYNKDMAIKYPKPGNPNPSVSLFVADLNSNASSNFSLWHNEPLAEPVVQNVLWVN 324

Qy         90 -------------------MIDAEGRII------------------------------- 98

Db        325 TSSVLVQFTNRNSTCITARLLDTELKSIHTVKTECLEEGWYEVQQSAKMFPLNNSLVWEN 384

Qy         99 ----------VDEVRRL---------------------------------VYFEGTKDSP 115

Db        385 WSDGYFDILALDDYNHLAFIPFNGSSPIYLTSGAWDVTDGPIHIDGDFGNVYFLATLKDS 444

Qy        116 LEHHLYVVSYVNPGEVTRLTDRGYS---HSCCISQHCDFFISKYSNQKNPHCVSLYKLSS 172

Db        445 TERHLYYVS-LDTLEIYGITDNGEDEGYYSTSFSPFGDFYVLNYHGPDVP----WQELRS 499

Qy        173 PEDDPTCKTKEFWATILDSAGPLPDYTPPEI----FSPESTTGFTLYGMLYKPHDLQPGK 228

Db        500 TKDKDYCLSLETNSRLKQQ---LSSITLPSVEYGKLTFNDTT-FNF--MERRPRNFDVNK 553

Qy        229 KYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLG------YVVVVIDNRGSCHRGLKF 282

Db        554 KYPVLFFAYGGPGSQQV------AKLFRVDFQAYLASHPDFEFIVVTLDGRGTGFNGNAF 607

Qy        283 EGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRV 342

Db        608 RYSVSRHLGEWESYDQGQAGKFWAD-LPFVDENHVGIWGWSYGGYLTLKTL-ETQDVFSY 665

Qy        343 AIAGAPVTLWIFYDTGYT 360

Db        666 GMAVAPVTDWRLYDSVYT 683


RESULT 7
S23752
dipeptidyl-peptidase IV (EC 3.4.14.5) alpha chain - mouse
N;Alternate names: CD26 alpha subunit; THAM alpha subunit
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S23752; A46465; A56030
R;Marguet, D.; Bernard, A.M.; Vivier, I.; Darmoul, D.; Naquet, P.; Pierres, M.
J. Biol. Chem. 267, 2200-2208, 1992
A;Title: cDNA cloning for mouse thymocyte-activating molecule. A multifunctional ecto
A;Reference number: S23752; MUID:92129288; PMID:1370813
A;Accession: S23752
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-760 <MAR>
A;Cross-references: EMBL:X58384
R;Vivier, I.; Marguet, D.; Naquet, P.; Bonicel, J.; Black, D.; Li, C.X.; Bernard, A.M
J. Immunol. 147, 447-454, 1991
A;Title: Evidence that thymocyte-activating molecule is mouse CD26 (dipeptidyl peptic
A;Reference number: A46465; MUID:91302787; PMID:1712807
A;Accession: A46465
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-20 <VIV>
A;Experimental source: M14.T thymoma cells, Swiss nu/nu
A;Note: sequence extracted from NCBI backbone (NCBIP:42236)
R;Bernard, A.M.; Mattei, M.G.; Pierres, M.; Marguet, D.
Biochemistry 33, 15204-15214, 1994
A;Title: Structure of the mouse dipeptidyl peptidase IV (CD26) gene.
A;Reference number: A56030; MUID:95092780; PMID:7999781
A;Accession: A56030
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 746-760 <BER>
A;Cross-references: GB:U12620
C;Genetics:
A;Gene: CD26
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein
F;213,223,315,514,679/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;624,702,734/Active site: Ser, Asp, His #status predicted

  Query Match             16.0%;  Score 309;  DB 1;  Length 760;
  Best Local Similarity   25.7%;  Pred. No. 4.3e-18;
  Matches  128;  Conservative   57;  Mismatches  154;  Indels  160;  Gaps   23;

Qy          1 EEDARSAGVATFVLQEE-FDRYSGYWWCPK------AETTPSGGKILRILYEENDESEVE 53

Db        184 EENVIYNGITDWVYEEEVFGAYSALWWSPNNTFLAYAQFNDTGVPLIEYSF-YSDES--- 239

Qy         54 IIHVTSPMLETRRADSFRYPKTGTANPKVTF---------------------------- 84

Db        240 --------LQYPKTVWIPYPKAGAVNPTVKFFIVNIDSLSSSSSAAPIQIPAPASVARGD 291

Qy         85 -KMSEIMIDAEGRIIVDEVRRLVYFEG-----------TKDSPLE-HHLYV-----VSYV 126

Db        292 HYLCDVVWATEERISLQWLRRIQNYSVMAICDYDKINLTWNCPSEQQHVEMSTTGWVGRF 351

Qy        127 NPGEVTRLTD-----------RGYSHSCCI---SQHCDF-----------------FISK 155

Db        352 RPAEPHFTSDGSSFYKIISDKDGYKHICHFPKDKKDCTFITKGAWEVISIEALTSDYLYY 411

Qy        156 YSNQ--KNPHCVSLYK-----------LSSPEDDPTCK------TKEFWATILDSAGP-L 195

Db        412 ISNQYKEMPGGRNLYKIQLTDHTNVKCLSCDLNPERCQYYAVSFSKEAKYYQLGCWGPGL 471

Qy        196 PDYT----------------------------PPEIFSF--ESTTGFTLYGMLYKPHDLQ 225

Db        472 PLYTLHRSTDHKELRVLEDNSALDRMLQDVQMPSKKLDFIVLNETRF-WYQMILPPH-FD 529
```

```
Qy        226 PGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLK 281
Db        530 KSKKYPLLLDVYAGPCSQ------KADASFRLNWATYLASTENIIVASFDGRGSGYQGDK 583

Qy        282 FEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFR 341
Db        584 IMHAINRRLGTLEVEDQIEAARQFV-KMGFVDSKRVAIWGWSYGGYVTSMVLGSGSGVFK 642

Qy        342 VAIAGAPVTLWIFYDTGYT 360
Db        643 CGIAVAPVSRWEYYDSVYT 661
```

RESULT 8
A39914
dipeptidyl-peptidase IV (EC 3.4.14.5), membrane-bound form precursor - rat
N;Alternate names: GP110; membrane glycoprotein 110K; OX-61
N;Contains: dipeptidyl-peptidase IV, soluble form
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A39914; A33315; B33315; A60730; A42203; S38949; A31781
R;Hong, W.; Doyle, D.
Proc. Natl. Acad. Sci. U.S.A. 84, 7962-7966, 1987
A;Title: cDNA cloning for a bile canaliculus domain-specific membrane glycoprotein of ra
A;Reference number: A39914; MUID:88068516; PMID:3479775
A;Accession: A39914
A;Molecule type: mRNA
A;Residues: 1-792 <HON>
A;Cross-references: GB:J02997; NID:g204463; PIDN:AAA41272.1; PID:g204464
R;Ogata, S.; Misumi, Y.; Ikehara, Y.
J. Biol. Chem. 264, 3596-3601, 1989
A;Title: Primary structure of rat liver dipeptidyl peptidase IV deduced from its cDNA an
A;Reference number: A33315; MUID:89123496; PMID:2563382
A;Accession: A33315
A;Molecule type: mRNA
A;Residues: 1-37,'A',39-182,'I',184-331,'T',333-351,'C',353-393,'V',395-561,'L',563-623,
A;Cross-references: GB:J04591; NID:g203973; PIDN:AAA41096.1; PID:g203974
A;Note: the authors translated the codon GCG for residue 38 as Arg, ACC for residue 332
A;Accession: B33315
A;Molecule type: protein
A;Residues: 1-20;35-54;427-443;505-509;511-520;530-538;593-600;602-608;618-627 <HO2>
R;McCaughan, G.W.; Wickson, J.E.; Creswick, P.F.; Gorrell, M.D.
Hepatology 11, 534-544, 1990
A;Title: Identification of the bile canalicular cell surface molecule GP110 as the ectop
quence.
A;Reference number: A60730; MUID:90228896; PMID:1970322
A;Accession: A60730
A;Molecule type: protein
A;Residues: 28-47,'XX',50-53,55-58 <MCC>
R;Ogata, S.; Misumi, Y.; Tsuji, E.; Takami, N.; Oda, K.; Ikehara, Y.
Biochemistry 31, 2582-2587, 1992
A;Title: Identification of the active site residues in dipeptidyl peptidase IV by affini
A;Reference number: A42203; MUID:92190188; PMID:1347701
A;Accession: A42203
A;Molecule type: protein
A;Residues: 'R',625-630,'X',632-648 <OG2>
R;Iwaki-Egawa, S.; Watanabe, Y.; Fujimoto, Y.
Biol. Chem. Hoppe-Seyler 374, 973-975, 1993
A;Title: N-terminal amino acid sequence of the 60-kDa protein of rat kidney dipeptidyl p
A;Reference number: S38949; MUID:94128239; PMID:7905271
A;Accession: S38949
A;Status: preliminary
A;Molecule type: protein
A;Residues: 281-302 <IWA>
R;Hong, W.; Doyle, D.
J. Biol. Chem. 263, 16892-16898, 1988
A;Title: Membrane orientation of rat gp110 as studied by in vitro translation.
A;Reference number: A31781; MUID:89034185; PMID:3182821
A;Accession: A31781
A;Molecule type: mRNA
A;Residues: 1-40 <HO3>
C;Comment: This protein is localized to the bile canaliculus, which is the apical domain
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptide hydrolase; glycoprotein; homodimer; liver; serine prote
F;1-792/Product: dipeptidyl-peptidase, membrane-bound form #status experimental <MATI
F;1-28/Domain: signal sequence #link MATS #status experimental <SIG>
F;1-6/Domain: intracellular #status predicted <INT>
F;7-28/Domain: transmembrane #status predicted <TMN>
F;29-792/Domain: extracellular #status predicted <EXT>
F;29-34/Domain: propeptide #link MATS #status experimental <PRO>
F;35-792/Product: dipeptidyl-peptidase, soluble form #status experimental <MATS>
F;83,90,148,217,227,319,521,686/Binding site: carbohydrate (Asn) (covalent) #status p
F;631/Active site: Ser #status experimental
F;709,741/Active site: Asp, His #status predicted

Query Match 15.1%; Score 291.5; DB 1; Length 792;
Best Local Similarity 24.1%; Pred. No. 1.4e-16;
Matches 119; Conservative 61; Mismatches 153; Indels 161; Gaps 22;

```
Qy          8 GVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLETRR 66
Db        195 GINDWVYEEEIFGAYSALWW------SPNGTFLAYAQFNDTGVPLIEYSFYSDESLQYPK 248

Qy         67 ADSFRYPKTGTANPKVTF----------------------------KMSEIMIDAEGR 96
Db        249 TVWIPYPKAGAVNPTVKFFIVNTDSLSSTTTTIPMQITAPASVTTGDHYLCDVAWVSEDR 308

Qy         97 IIVDEVRRLVYF--------------------EGTKDS---------PLEHHLYVVSYV 126
Db        309 ISLQWLRRIQNYSVMAICDYDKTNLVWNCPTTQEHIETSATGWVGRFRPAEPH-----FT 363

Qy        127 NPGE--VTRLTDR-GYSHSCCI------SQHCDF----------------FISKYSNQ- 159
Db        364 SDGSSFYKIVSDKDGYKHICQFQKDRKPEQDCTFITKGAWEVISIEALTSDYLYYISNEY 423

Qy        160 -KNPHCVSLYK-----------LSSPEDDPTCK------TKEFWATILDSAGP-LPDYT- 199
Db        424 KEMPGGRNLYKIQLTDHTNKKCLSCDLNPERCQYYSVSLSKEAKYYQLGCRGPGLPLYTL 483

Qy        200 -----PPEIFSFESTTGFT----------------------LYGMLYKPHDLQPGKKYP 231
Db        484 HRSTDQKELRVLEDNSALDKMLQDVQMPSKKLDFIVLNETRFWYQMILPPH-FDKSKKYP 542

Qy        232 TVLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLKFEGAFK 287
Db        543 LLIDVYAGPCSQ------KADAAFRFNWATYLASTENIIVASFDGRGSGYQGDKIMHAIN 596

Qy        288 YKMGQIEIDDQVEGL-QYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAG 346
Db        597 KRLGTLEVEDQIEAARQFL--KMGFVDSKQVAIWGWSYGGYVTSMVLGSGSGVFKCGIAV 654

Qy        347 APVTLWIFYDTGYT 360
Db        655 APVSRWEYYDSVYT 668
```

RESULT 9
A30107
dipeptidyl aminopeptidase B (EC 3.4.14.-) - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YHR028c
C;Species: Saccharomyces cerevisiae
C;Date: 07-Jun-1990 #sequence_revision 30-May-1997 #text_change 23-Mar-2001
C;Accession: S46780; A30107
R;Du, Z.
submitted to the EMBL Data Library, June 1994
A;Description: The sequence of S. cerevisiae cosmid 8082.
A;Reference number: S46773
A;Accession: S46780
A;Molecule type: DNA
A;Residues: 1-818 <DUZ>
A;Cross-references: EMBL:U10399; NID:g500689; PIDN:AAB68879.1; PID:g500698; GSPDB:GN0
R;Roberts, C.J.; Pohlig, G.; Rothman, J.H.; Stevens, T.H.
J. Cell Biol. 108, 1363-1373, 1989
A;Title: Structure, biosynthesis, and localization of dipeptidyl aminopeptidase B, ar
A;Reference number: A30107; MUID:89174971; PMID:2647766
A;Accession: A30107

```
A;Molecule type: DNA
A;Residues: 1-82,'H',84-124,'N',126-181,'LRRLET',189-199,'N',201-365,'DFKRGKERKF',376-57
A;Cross-references: EMBL:X15484
A;Note: the authors translated the codon ACC for residue 572 as Asn
C;Genetics:
A;Gene: SGD:DAP2; STE13; MIPS:YHR028c
A;Cross-references: SGD:S0001070; MIPS:YHR028c
A;Map position: 8R
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein; yeast vacu
F;30-45/Domain: transmembrane #status predicted <TMM>
F;63,79,110,139,392,421/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             14.5%;  Score 280.5;  DB 1;  Length 818;
  Best Local Similarity   22.0%;  Pred. No. 1.3e-15;
  Matches  114;  Conservative   50;  Mismatches  155;  Indels  199;  Gaps   19;

Qy          8 GVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEV-EIIHVTSPMLETR 65
              |   :| :|| |:     || |  :          : : : ||||| | |    |
Db        234 GKPDWVYEEEVFEDDKAAWWSPTGD---------YLAFLKIDESEVGEFI---IPYYVQD 281

Qy         66 RAD------SFRYPKTGTANP---------------------------------------- 80
                |       | :|||:|| ||
Db        282 EKDIYPEMRSIKYPKSGTPNPHAELWVYSMKDGTSFHPRISGNKKDGSLLITEVTWVGNG 341

Qy         81 ----KVTFKMSEI----MID---------------------------------------- 92
                  | | : |:|    :||
Db        342 NVLVKTTDRSSDILTVFLIDTIAKTSNVVRNESSNGGWWEITHNTLFIPANETFDRPHNG 401

Qy         93 ------------------------------------AEGRIIVDEVRRLVYFEGTKDSPLEH 118
                                                  | :  | :   :||  |: |  |
Db        402 YVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTRKSSTER 461

Qy        119 HLYVVSYVNPGEVTRLTDRGYS--HSCCISQHCDFFISKYSNQKNPH------------- 163
              |:| :   :| |:  :||       :    |    | :  |    | |:
Db        462 HVYYIDLRSPNEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEK 521

Qy        164 C-------VSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYG 216
              |       |||  |   |              :|     | ||  |      |   |   :|
Db        522 CDKGNVLGKSLYHLEKNE-------------VLTKI--LEDYAVPRKSFRELNLGKDEFG 566

Qy        217 --MLYKPHDLQP-------GKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN----TLASL 263
                :|   ::: |           ||   | ||||  | |      || | :       : |
Db        567 KDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQV------VKTFSVGFNEVVASQL 620

Qy        264 GYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWS 323
                :|||:| ||:  :|  |    : ::|  |  ||:       |   |:|  :: : |||
Db        621 NAIVVVVDGRGTGFKGQDFRSLVRDRLGDYEARDQISAASLYGS-LTFVDPQKISLFGWS 679

Qy        324 YGGYLSLMALMQRSD-IFRVAIAGAPVTLWIFYDTGYT 360
              |||||:|  | :     |:  :: |||| | |||: ||
Db        680 YGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYT 717


RESULT 10
A49737
dipeptidyl aminopeptidase (EC 3.4.14.-) - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein O5045; protein YOR219c; protein YOR50-9
C;Species: Saccharomyces cerevisiae
C;Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 29-Oct-1999
C;Accession: A49737; S45451; S60946; S67112; S71721
R;Santa Anna-A, S.; Herskowitz, I.
submitted to the Protein Sequence Database, July 1993
A;Reference number: A49737
A;Accession: A49737
A;Molecule type: DNA
A;Residues: 1-931 <SAN>
A;Cross-references: GB:L21944; NID:g347196; PIDN:AAA35119.1; PID:g347197
R;Anna-Arriola, S.S.; Herskowitz, I.
Yeast 10, 801-810, 1994
A;Title: Isolation and DNA sequence of the STE13 gene encoding dipeptidyl aminopeptidase
A;Reference number: S45451; MUID:95066382; PMID:7975897
A;Accession: S45451
A;Molecule type: DNA
A;Residues: 1-931 <ANN>
A;Cross-references: EMBL:L21944; NID:g347196; PIDN:AAA35119.1; PID:g347197
R;Galisson, F.; Dujon, B.
submitted to the EMBL Data Library, October 1995
A;Description: Sequence and analysis of a 33 kb fragment from the right arm of chromo
A;Reference number: S60938
A;Accession: S60946
A;Molecule type: DNA
A;Residues: 1-931 <GAL>
A;Cross-references: EMBL:X92441; NID:g1050762; PIDN:CAA63182.1; PID:g1050771
R;Boyer, J.; Fairhead, C.; Gaillon, L.; Galisson, F.; Michaux, G.; Thierry, A.; Dujon
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67104
A;Accession: S67112
A;Molecule type: DNA
A;Residues: 1-931 <BOY>
A;Cross-references: EMBL:Z75127; NID:g1420507; PIDN:CAA99437.1; PID:e252398; PID:g142
A;Experimental source: strain S288C
R;Galisson, F.; Dujon, B.
Yeast 12, 877-885, 1996
A;Title: Sequence and analysis of a 33 kb fragment from the right arm of chromosome X
A;Reference number: S71713; MUID:96437977; PMID:8840505
A;Accession: S71721
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-931 <GAW>
A;Cross-references: EMBL:X92441; NID:g1050762; PIDN:CAA63182.1; PID:g1050771
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995
C;Genetics:
A;Gene: SGD:STE13; YCI1
A;Cross-references: SGD:S0005745; MIPS:YOR219c
A;Map position: 15R
C;Function:
A;Description: involved in processing of alpha-factor prepropheromone
C;Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein
F;113-150/Domain: transmembrane #status predicted <TMM>
F;377/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;785,863,896/Active site: Ser, Asp, His #status predicted

  Query Match             12.9%;  Score 250;  DB 2;  Length 931;
  Best Local Similarity   24.3%;  Pred. No. 6e-13;
  Matches  101;  Conservative   58;  Mismatches  162;  Indels   94;  Gaps   15;

Qy         15 QEEFDRYSGYWWCP---KAETTPSGGKILRILYEENDESEVEIIHVTS------------ 59
              |::   |:| |   |    : | |    ||| :   :    |::  :  |:
Db        434 QKDSILYNGKWISPDTFRFEITDRNSKILDVKVYDIPSSQMLTVRNTNSNLFNGWIEKTK 493

Qy         60 ------PMLETRR---------ADS------FRYPKTGTANPKVTFKMSEIMIDAEGRII 98
                    |  | :|         |||      | || |  |   :        :   | :
Db        494 DILSIPPKPELKRMDYGYIDIHADSRGFSHLFYYP-TVFAKEPIQLTKGNWEVTGNGIVG 552

Qy         99 VDEVRRLVYFEGTKDSPLEHHLYVVSYV------------NPGEVTRLTDRGYSHSCCIS 146
               :    ::|   :   :  ||| :|               ||      :|:   :   :|
Db        553 YEYETDTIFFTANEIGVMSQHLYSISLTDSTTQNTFQSLQNP------SDKYDFYDFELS 606

Qy        147 QHCDFFISKYSNQKNP---------------HCVSLYKLSSPEDDPTCKTKEFWATILD 190
                : |||          |               |  |: :|:  |     | ||
Db        607 SSARYAISKKLGPDTPIKVAGPLTRVLNVAEIHDDSILQLTKDE-----KFKE------- 654

Qy        191 SAGPLPDYTPPEIFSFESTT---GFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
                  : :|   | | |:::     |  :  :  || :| | |||| :: |||||  |
Db        655 ---KIKNYDLP-ITSYKTMVLDDGVEINYIEIKPANLNPKKKYPILVNIYGGPGSQTFTT 710

Qy        248 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVE-GLQYLA 306
                |     |    :: |  :|: |: ||:  :|  |     : |:|  |  |  |   :::
Db        711 --KSSLAFEQAVVSGLDVIVLQIEPRGTGGKGWSFRSWAREKLGYWEPRDITEVTKKFIQ 768

Qy        307 SRYDFIDLDRVGIHGWSYGGYLSLMAL-MQRSDIFRVAIAGAPVTLWIFYDTGYT 360
                 ||  :: | ||||||: ||  : :    | |: |:| |||| |  ||: ||
Db        769 RNSQHIDESKIAIWGWSYGGFTSLKTVELDNGDTFKYAMAVAPVTNWTLYDSVYT 823
```

RESULT 11
F82858
dipeptidyl-peptidase XF0015 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: F82858
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequer
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: F82858
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-795 <SIM>
A;Cross-references: GB:AE003856; GB:AE003849; NID:g9104770; PIDN:AAF82828.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0015

Query Match 12.7%; Score 245; DB 2; Length 795;
Best Local Similarity 27.5%; Pred. No. 1.3e-12;
Matches 98; Conservative 49; Mismatches 140; Indels 70; Gaps 16;

```
Qy         21 YSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANP 80
              :: : | |   |       | :| ||:  | :  |             .     :|||
Db        396 FNDFGWLPDNRT-------LWLLSEESGYSHLYTIQ----------------ETGTPKQ 431

Qy         81 KVTF---KMSEIMIDAEGRIIVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPG--EVTRL- 134
              : ||   ::|     |:|         | |||             | :   ||  || :|
Db        432 R-TFGHWEVSAPRFSADG--------RSVYF-------------VCNPTWPGGYEVCKLD 469

Qy        135 TDRGYSHSCCISQHC-DFFISKYSNQKNPHCVSLYKLS-SPEDDPTCKTKEFWATIL--- 189
               | |:  :    |   || :|     |     : ||  |  |       |  | :|
Db        470 LDSGHFDTLTALQGVEDFDLSPDGRQ----ILVLYSGSYLPPQLAVISTNGGQARVLTDT 525

Qy        190 -DSAGPLPDYTPPEIFSFESTTGF-TLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
               |      :  ||     |  |   ::|  | |  ||| |:|| |:|::|   :| |:
Db        526 RTSTFKAKQWIQPEYVQIPSKHGAGVIWGKYYGPASLQPEKRYPIVMFVHGAGYLQNVSA 585

Qy        248 RFKGVKYFRLNTLASL----GYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQ 303
              |:    |||     :|    ||:|: :| | |   |  :  |     ||  |::| ::||
Db        586 RYP--SYFREQMFHTLLVQKGYIVLDLDYRASAGHGRDWRTAIYRNMGHPELEDYLDGLD 643

Qy        304 YLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              :| :     |  | |::| ||||::: ||: :    |:   |  ||  |: |:  ||
Db        644 WLVAHKQG-DRRRAGMYGGSYGGFMTYMAMFRAPGTFKAGAALRPVGDWMQYNHEYT 699
```

RESULT 12
T19514
hypothetical protein C27C12.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C;Accession: T19514
R;Thomas, K.
submitted to the EMBL Data Library, March 1996
A;Reference number: Z19134
A;Accession: T19514
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-829 <WIL>
A;Cross-references: EMBL:Z69883; PIDN:CAA93743.1; GSPDB:GN00028; CESP:C27C12.7
A;Experimental source: clone C27C12
C;Genetics:
A;Gene: CESP:C27C12.7
A;Map position: X
A;Introns: 4/3; 51/1; 123/3; 166/3; 188/2; 279/2; 392/3; 501/3; 553/1; 583/1; 606/2;
C;Superfamily: dipeptidyl-peptidase IV

Query Match 12.4%; Score 240; DB 2; Length 829;
Best Local Similarity 21.1%; Pred. No. 3.7e-12;
Matches 107; Conservative 77; Mismatches 144; Indels 178; Gaps 20;

```
Qy          8 GVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLETRR 66
              |:| :: :||  :  |  ||       :|||  :  : ::: : : : :   |
Db        249 GLANWLYEEEILEASSAVWW------SPSGRYVSYLRFDDREVNRIFLPKYTDD------ 296

Qy         67 ADSF------RYPKTGTANPKVTFKMSEIMIDAEGRIIVDEV------------------ 102
               ||:       ||| |  |  :   ::: : |:|   ||:
Db        297 -DSYVEYFELPYPKAGVQNNTL---VTQYIWDSENHKIVETAPPNELSAANGDYYVLTNK 352

Qy        103 -------------RRLVYFEGTKDSPLEHHLYV-------------------------- 122
                          |||     :|   ::|:|
Db        353 WITMPRNGSDLGEERLVTVWANRD---QNHVYFSLCNEQDCVMALSFQFSIDNRQLWVSP 409

Qy        123 --VSYVNPGEVTRLT------DRG--YSH-------------SCCISQHCDFFI----- 153
                 |  | | |   ||       | |  |:|              :  | :: |  :
Db        410 KDVRGVFPTETGFLTVLPHKHDDGNIYNHVAHVELDGTGTGKITKWIGENFDVILVLGYS 469

Qy        154 -------------------------SKYSNQKNPHCVSLYK----------LSSPEDDPT 178
                                       : |||:|     :| |           | |   |||
Db        470 SKIDALTFSAYGDGVGEFSTYIVREAMYSNKK----TTLQKVTDQFEDCKTLGSQSADPT 525

Qy        179 -------CK-----TKEFWATILDSA----------GPLPDYTPPEIF-SFESTTGFTLY 215
                    |:     |: :   ::|:             :|   |  |   :  :|   :
Db        526 GQRIVVQCEKPFDNTRLYLVDVVDTTKKIMLEGGTKAVIPFDVPNMKFGKLKLPSGIDGH 585

Qy        216 GMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGS 275
              |:  | :|  | | | :| |||||  : |  :          :: :   : |    || ||:
Db        586 YMMLTPANLLDGAKIPLLLDIYGGPDSKQVFQKTPTAHAIQIVSQYDIAY--ARIDVRGT 643

Qy        276 CHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYL-SLMALM 334
               ||   : |   |:|  |: | :: ::   : : ||| ||: : ||||||:| | :|:
Db        644 GGRGWDVKEAVYRKLGDAEVVDTLDMIRAFINTFGFIDEDRIAVMGWSYGGFLTSKIAIK 703

Qy        335 QRSDIFRVAIAGAPVTLWIFYDTGYT 360
               : :: : ||: |||| : :||: ||
Db        704 DQGELVKCAISIAPVTDFKYYDSAYT 729
```

RESULT 13
E87495
prolyl oligopeptidase family protein [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: E87495
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: E87495
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-683 <STO>
A;Cross-references: GB:AE005673; NID:g13423453; PIDN:AAK23961.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1986

Query Match 11.7%; Score 226.5; DB 2; Length 683;

```
  Best Local Similarity   32.2%;  Pred. No. 4e-11;
  Matches   66;  Conservative   28;  Mismatches   94;  Indels   17;  Gaps    6;

Qy        126 VNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFW 185
              |  |:|: ||  |:  :   :       |  |    : |:|     |: |   |   |  |
Db        356 VKTGKVSALTGEGHVTAFDVGPSGIVFAS--DSLKSPS--ELFLL--PAKGPAVKV---- 405

Qy        186 ATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLV 245
              |::    |      :   || |||:     |::| | || :  | ||||     |:||||
Db        406 ASVSSEALKDVAWGEPEQFSFKGWNDETVHGFLVKPANFDPAKKYPVAFLIHGGPQ---- 461

Qy        246 NNRFKGVKYFRLN--TLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQ 303
                 |      :| |    |: || ||:|| ||    |   |  :       |    ::|  :|
Db        462 -GSFSNAWSYRWNPQVYANAGYAVVMIDFHGSTGYGQAFTDSISRHWGDRPLEDLQKGWS 520

Qy        304 YLASRYDFIDLDRVGIHGWSYGGYL 328
              :: |:| |:| ||      | |||||:
Db        521 FVLSKYGFLDGDRACALGASYGGYM 545

RESULT 14
A41793
dipeptidyl aminopeptidase-like protein (EC 3.4.14.-) short form, DPPX-S - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C;Accession: A41793
R;Wada, K.; Yokotani, N.; Hunter, C.; Doi, K.; Wenthold, R.J.; Shimasaki, S.
Proc. Natl. Acad. Sci. U.S.A. 89, 197-201, 1992
A;Title: Differential expression of two distinct forms of mRNA encoding members of a dip
A;Reference number: A41793; MUID:92108018; PMID:1729689
A;Accession: A41793
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-803 <WAD>
A;Cross-references: GB:M76429; NID:g408719; PIDN:AAC41623.1; PID:g408720
A;Note: sequence extracted from NCBI backbone (NCBIP:75138)
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptide hydrolase; glycoprotein
F;257,342/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             11.4%;  Score 221;  DB 2;  Length 803;
  Best Local Similarity   21.0%;  Pred. No. 1.5e-10;
  Matches  103;  Conservative   58;  Mismatches  169;  Indels  160;  Gaps   16;

Qy          8 GVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV---TSPMLE 63
              |:: :: :||    :  :||  |           |:  |    ::|  | : :   |    :
Db        225 GLSDWLYEEEILKTHIAHWWSPDG---------TRLAYATINDSRVPVMELPTYTGSVYP 275

Qy         64 TRRADSFRYPKTGTANPKVTF-------------------KMSEIMI----------- 91
               |   |   : ||| |  || ::                  :| |  |
Db        276 T--AKPYHYPKAGCENPSISLHVIGLNGPTHDLEMTPPDDPRMREYYITMVKWATSTKVA 333

Qy         92 ----------------DA---------------------------EGR----------- 96
                              ||                            :||
Db        334 VNWLSRAQNVSILTLCDATTGVCTKKHEDESEAWLHRQNEEPVFSKDGRKFFFVRAIPQG 393

Qy         97 ----------------------------------IIVDEVRRLVYFEGTKDSPLEHHLY 121
                                                : || |  :|| |:| |    ||
Db        394 GQGKFYHITVSSSQPNSSNDNIQSITSGDWDVTKILSYDEKRSQIYFLSTEDLPRRRQLY 453

Qy        122 VVSYVNPGEVTR------LTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPED 175
               | |  |   |       | |    |    |   |||: |     |  ||:: :   :
Db        454 SASTV--GSFNRQCLSCDLVDNCTYFSASFSPGADFFLLKCEGPGVP-TVSVHNTTDKKK 510

Qy        176 DPTCKTKE-FWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVL 234
                 :| |      | |    |  :|   |       |  : |   : ||          || :|
Db        511 MFDLETNEHVQKAISDRQMPKVEYRKIE------TDDYNLPIQILKPATFTDTAHYPLLL 564

Qy        235 FIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIE 294
                : |  |  | |  :|       :     ::| | |||  ||||  :| :    : ::| :|
Db        565 VVDGTPGSQSVAEKF--AVTWETVMVSSHGAVVVKCDGRGSGFQGTRLLHEVRRRLGSLE 622

Qy        295 IDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSD----IFRVAIAGAPVT 350
```

```
                ||:| :: :  :  :|| || : |  |||||  |  :  |     :|    | :|:|
Db        623 EKDQMEAVRVML-KEPYIDKTRVAVFGKDYGGYLSTYLLPAKGDGQAPVFSCGSALSPIT 681

Qy        351 LWIFYDTGYT 360
                :  |  : ::
Db        682 DFKLYASAFS 691

RESULT 15
B82580
alanyl dipeptidyl peptidase XF2260 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: B82580
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: B82580
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-709 <SIM>
A;Cross-references: GB:AE004038; GB:AE003849; NID:g9107408; PIDN:AAF85059.1; GSPDB:GN
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF2260

  Query Match             11.4%;  Score 220;  DB 2;  Length 709;
  Best Local Similarity   28.5%;  Pred. No. 1.5e-10;
  Matches   73;  Conservative   32;  Mismatches   99;  Indels   52;  Gaps    9;

Qy         88 EIMIDAEGRIIVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQ 147
              || : |:|: :          |      |     || |: |   :  |:| :      |  |: ::
Db        358 EIALSADGKAL--------YVNA--DDHGEHPLFKVD-IASGKVEKWVGEGSVHAPVLAG 406

Qy        148 HCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPP------ 201
                    |  :  :|         | |  :         : :  |   : |||   |
Db        407 ------GKLAFARN-------SLKSAD---------QIFVTDAVARGPLQAITSATGEVLQ 445

Qy        202 -------EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKY 254
                     | |||:       |:|| : ||:| ||||||     |:||||          |
Db        446 QVRLGDFEQFSFKGWNDETVYGYVVKPYDYQPGKKYPVAFLIHGGPQ-----GSFGNSWG 500

Qy        255 FRLN--TLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFI 312
               :| |  | |    || ||:||  ||     |   |    |       |   ::|  :|        :| |:
Db        501 YRWNPQTYAGQGYAVVMIDFHGSTGYGQAFTDAISQHWGDRPLEDLQKGWAAAQQQYPFL 560

Qy        313 DLDRVGIHGWSYGGYL 328
               : |:     |  |||||:
Db        561 NGDKACALGASYGGYM 576

Search completed: October 15, 2003, 17:14:52
Job time : 11.4596 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 15, 2003, 16:57:30 ; Search time 5.71145 Seconds
(without alignments)
2964.152 Million cell updates/sec

Title: US-10-070-464-7
Perfect score: 1933
Sequence: 1 EEDARSAGVATFVLQEEFDR.....RVAIAGAPVTLWIFYDTGYT 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	320	16.6	765	1	DPP4_FELCA	Q9n2i7 felis silve
2	319	16.5	765	1	DPP4_BOVIN	P81425 bos taurus
3	315	16.3	766	1	DPP4_HUMAN	P27487 homo sapien
4	312.5	16.2	793	1	YEA8_SCHPO	O14073 schizosacch
5	309	16.0	760	1	DPP4_MOUSE	P28843 mus musculu
6	297.5	15.4	767	1	DPP4_RAT	P14740 rattus norv
7	280.5	14.5	818	1	DAP2_YEAST	P18962 saccharomyc
8	259	13.4	761	1	SEPR_MOUSE	P97321 mus musculu
9	257	13.3	760	1	SEPR_HUMAN	Q12884 homo sapien
10	250	12.9	931	1	ST13_YEAST	P33894 saccharomyc
11	221	11.4	863	1	DPP6_BOVIN	P42659 bos taurus
12	212	11.0	859	1	DPP6_RAT	P46101 rattus norv
13	211	10.9	804	1	DPP6_MOUSE	Q9z218 mus musculu
14	203	10.5	865	1	DPP6_HUMAN	P42658 homo sapien
15	197.5	10.2	853	1	YDZF_SCHPO	Q9p7e9 schizosacch
16	196	10.1	657	1	YUXL_BACSU	P39839 bacillus su
17	134.5	7.0	732	1	ACPH_HUMAN	P13798 homo sapien
18	131.5	6.8	732	1	ACPH_RAT	P13676 rattus norv
19	126.5	6.5	689	1	PPCE_AERHY	Q06903 aeromonas h
20	126.5	6.5	732	1	ACPH_PIG	P19205 sus scrofa
21	118	6.1	436	1	PAFA_CAVPO	P70683 c platelet-
22	112.5	5.8	835	1	NLG2_HUMAN	Q8nfz4 homo sapien
23	110.5	5.7	836	1	NLG2_RAT	Q62888 rattus norv
24	106	5.5	705	1	PPCF_FLAME	P27195 flavobacter
25	106	5.5	710	1	PPCE_PIG	P23687 sus scrofa
26	105	5.4	722	1	Y174_RICPR	P81171 rickettsia
27	103.5	5.4	705	1	PPCE_FLAME	P27028 flavobacter
28	102.5	5.3	489	1	PNBA_BACSU	P37967 bacillus su
29	100	5.2	710	1	PPCE_BOVIN	Q9xta2 bos taurus
30	100	5.2	710	1	PPCE_MOUSE	Q9qur6 mus musculu
31	98.5	5.1	444	1	PAFA_CANFA	Q28262 c platelet-
32	98	5.1	705	1	Y4SO_RHISN	P55656 rhizobium s
33	96	5.0	710	1	PPCE_HUMAN	P48147 homo sapien
34	95	4.9	740	1	YL31_CAEEL	P34422 caenorhabdi
35	95	4.9	848	1	NLG3_RAT	Q62889 rattus norv
36	94.5	4.9	840	1	NLG1_HUMAN	Q8n2q7 homo sapien
37	94.5	4.9	3354	1	CADN_HUMAN	Q9h251 homo sapien
38	94	4.9	440	1	PAFA_MOUSE	Q60963 m platelet-
39	94	4.9	848	1	NLG3_HUMAN	Q9nz94 homo sapien
40	93.5	4.8	419	1	CARV_CANAL	P10977 candida alb
41	93	4.8	367	1	DUS1_MOUSE	P28563 mus musculu
42	92.5	4.8	563	1	LIP2_GEOCN	P22394 geotrichum
43	92.5	4.8	843	1	NLG1_RAT	Q62765 rattus norv
44	92	4.8	816	1	NLG4_HUMAN	Q8n0w4 homo sapien
45	92	4.8	825	1	NLG3_MOUSE	Q8bym5 mus musculu

ALIGNMENTS

RESULT 1
DPP4_FELCA
ID DPP4_FELCA STANDARD; PRT; 765 AA.
AC Q9N2I7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26).
GN DPP4 OR CD26.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood;
RX MEDLINE=20094000; PubMed=10630304;
RA Nishimura Y., Miyazawa T., Ikeda Y., Izumiya Y., Nakamura K., Sato E.,
RA Mikami T., Takahashi E.;
RT "Molecular cloning and sequencing of a cDNA encoding the feline T-cell
RT activation antigen CD26 homologue.";
RL Immunogenetics 50:366-368(1999).
CC -!- FUNCTION: Removes N-terminal dipeptides sequentially from
CC polypeptides having unsubstituted N-termini provided that the
CC penultimate residue is proline (By similarity).
CC -!- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC Xcc is neither Pro nor hydroxyproline.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC a soluble form (By similarity).
CC -!- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC by proteolytic processing (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family S9B.
CC ---

CC ---
DR EMBL; AB023952; BAA92344.1; -.
DR MEROPS; S09.003; -.
DR InterPro; IPR002469; DPPIV_N_term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002471; Prol_endopep_ser.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00930; DPPIV_N_term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW Transmembrane; Glycoprotein; Signal-anchor.
FT CHAIN 1 765 DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM

```
FT                                (MDPP).
FT   CHAIN        38    765       DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT                                (SDPP) (BY SIMILARITY).
FT   DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     29       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       30    765       EXTRACELLULAR (POTENTIAL).
FT   ACT_SITE    629    629       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    707    707       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    739    739       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     84     84       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     91     91       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    149    149       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    178    178       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    228    228       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    280    280       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    320    320       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    330    330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    331    331       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    519    519       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    684    684       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   765 AA;  88213 MW;  3EFCE98A22B175D9 CRC64;

  Query Match             16.6%;  Score 320;  DB 1;  Length 765;
  Best Local Similarity   25.1%;  Pred. No. 1.1e-19;
  Matches  125;  Conservative   59;  Mismatches  156;  Indels  158;  Gaps   22;

Qy          1 EEDARSAGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH--- 56
              ||:|   |:| :| :|| |  ||  || ||            : | : ::::| :|
Db        189 EENAIYNGIADWVYEEEIFSAYSALWWSPKGTF---------LAYAQFNDTQVPLIEYSF 239

Qy         57 VTSPMLETRRADSFRYPKTGTANPKV-----------------------TFKM 86
               :   |:        ||| | ||| |                       : :
Db        240 YSDESLQYPMTMRIPYPKAGAANPTVKLFVIKTDNLNPNTNATSVEITPPAAMLTGDYYL 299

Qy         87 SEIMIDAEGRIIVDEVRRLV---------YFEGTK---DSPLEHHLYV-----VSYVNPG 129
               ::    | || :  :||:         |   |     | : |: :     |    |
Db        300 CDVTWANEERISLQWLRRIQNYSVMDIRDYNNSTGKWISSAAQEHIEMSTTGWVGRFRPA 359

Qy        130 EVTRLTD-----------RGYSHSCCI---SQHCDF-----------------FISKYSN 158
              |    :|           || | |      : | |                 ::   ||
Db        360 EPHFTSDGRNFYKIISNEDGYKHICRFQIDKKDCTFITKGAWEVIGIEALTTDYLYYISN 419

Qy        159 QKN--PHCVSLYKLS--------------SPEDDPTCK------TKEFWATILDSAGP-L 195
              :    |   :|||:                ||    |:      :||     |  :|| |
Db        420 EYKGMPGGRNLYKIQLNDYTKVACLSCELKPE---RCQYYSVSFSKEAKYYQLRCSGPGL 476

Qy        196 PDYT------PPEIFSFESTTGFT----------------------LYGMLYKPHDLQP 226
              | ||        |:   |  :                        | |:  ||
Db        477 PLYTLHRSSNDEELRVLEDNSALDKMLQEVQMPSKKLDFIILNETKFWYQMILPPH-FDT 535

Qy        227 GKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLKF 282
              |||| :: :| ||  |      |    ||||    |||   ::|   | ||| ::| |
Db        536 SKKYPLLIDVYAGPCSQ------KADAIFRLNWATYLASTENIIVASFDGRGSGYQGDKI 589

Qy        283 EGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRV 342
              |   ::|  |::||:|  :   |:  |:|  |: | |||||||:: | |   | :|:
Db        590 MHAVNRRLGTFEVEDQIEAARQF-SKMGFVDDKRIAIWGWSYGGYVTSMVLGAGSGVFKC 648

Qy        343 AIAGAPVTLWIFYDTGYT 360
              || |||: | :||: ||
Db        649 GIAVAPVSRWEYYDSVYT 666


RESULT 2
DPP4_BOVIN
ID   DPP4_BOVIN     STANDARD;      PRT;   765 AA.
AC   P81425; Q8WMG8;
DT   15-DEC-1998 (Rel. 37, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE   antigen CD26) (Adenosine deaminase complexing protein) (ADCP-I)
DE   (Activation molecule 3) (ACT3) (WC10).
GN   DPP4 OR CD26.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymphocytes;
RX   MEDLINE=22067734; PubMed=12073152;
RA   Lee S.-U., Park Y.-H., Davis W.C., Hamilton M.J., Naessens J.,
RA   Bohach G.A.;
RT   "Molecular characterization of bovine CD26 upregulated by a
RT   staphylococcal superantigen.";
RL   Immunogenetics 54:216-220(2002).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.
RC   TISSUE=Thymus;
RX   MEDLINE=22021197; PubMed=11981836;
RA   Gliddon D.R., Howard C.J.;
RT   "CD26 is expressed on a restricted subpopulation of dendritic cells in
RT   vivo.";
RL   Eur. J. Immunol. 32:1472-1481(2002).
RN   [3]
RP   SEQUENCE OF 1-24.
RC   TISSUE=T-cell;
RX   MEDLINE=21482004; PubMed=11598101;
RA   Lee S.-U., Ferens W., Davis W.C., Hamilton M.J., Park Y.-H., Fox L.K.,
RA   Naessens J., Bohach G.A.;
RT   "Identity of activation molecule 3 on superantigen-stimulated bovine
RT   cells is CD26.";
RL   Infect. Immun. 69:7190-7193(2001).
RN   [4]
RP   SEQUENCE OF 537-546.
RC   TISSUE=Kidney;
RX   MEDLINE=98293306; PubMed=9629661;
RA   Ben-Shooshan I., Parola A.H.;
RT   "The CP-I subunit of adenosine deaminase complexing protein from calf
RT   kidney is identical to human, mouse, and rat dipeptidyl peptidase
RT   IV.";
RL   Comp. Biochem. Physiol. 119B:289-292(1998).
CC   -!- FUNCTION: Removes N-terminal dipeptides sequentially from
CC       polypeptides having unsubstituted N-termini provided that the
CC       penultimate residue is proline. Binds and regulates the activity
CC       of ADA.
CC   -!- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC       Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC       Xcc is neither Pro nor hydroxyproline.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC       a soluble form (By similarity).
CC   -!- TISSUE SPECIFICITY: Intestinal epithelium, dendritic cells and
CC       several immune system tissues.
CC   -!- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC       by proteolytic processing (By similarity).
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF461806; AAL67836.1; -.
DR   EMBL; AY056834; AAL23628.1; -.
DR   MEROPS; S09.003; -.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
```

```
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;.
KW   Transmembrane; Glycoprotein; Signal-anchor.
FT   CHAIN         1    765       DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT                                (MDPP).
FT   CHAIN        38    765       DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT                                (SDPP) (BY SIMILARITY).
FT   DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     29       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       30    765       EXTRACELLULAR (POTENTIAL).
FT   ACT_SITE    629    629       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    707    707       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    739    739       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     84     84       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     91     91       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    149    149       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    218    218       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    228    228       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    271    271       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    280    280       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    320    320       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    392    392       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    495    495       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    684    684       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   765 AA;  88369 MW;  E32165421F43E116 CRC64;

  Query Match           16.5%;  Score 319;  DB 1;  Length 765;
  Best Local Similarity   24.5%;  Pred. No. 1.3e-19;
  Matches  125;  Conservative   56;  Mismatches  147;  Indels  182;  Gaps   22;

Qy          1 EEDARSAGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH--- 56
              ::|    |:  :| :|| |  ||  || | :           : | : :::|| :|
Db        189 KKDVIYNGITDWVYEEEVFSAYSALWWSPNSTF---------LAYAQFNDTEVPLIEYSF 239

Qy         57 VTSPMLETRRADSFRYPKTGTANPKVTFKMSEIM-----IDAEGRIIVDEVRRLVYFEGT 111
               :   |:  :     ||| |  || : | :  |      |:|  : ||     |:
Db        240 YSDESLQYPKTVKIPYPKAGAVNPTIKFFVVNISSLSPNINATSQQIVPPGSVLIG---- 295

Qy        112 KDSPLEHHLYVVSYVNPGEVT-----RLTDRGYSHSCCISQHCDF---:---FISKYSNQ- 159
                    :|:|  |::|    ::     |:  : ||    |   ||:     :||    |
Db        296 -----DHYLCDVTWVTEERISLQWLRRI--QNYS----IMDICDYDRSTGRWISSVGRQH 344

Qy        160 --------------KNPHCV----SLYKLSSPE-----------DDPTCK--TKEFWATI 188
                           ||      | ||: | |           |   |   || | |
Db        345 IEISTTGWVGRFRPAEPHFTSDGNSFYKIISNEEGYKHICHFQTDKRNCTFITKGAWEVI 404

Qy        189 ------------------------------------------------------------ 188

Db        405 GIEALTSDYLYYISNEYKGMPGARNLYKIQLNDYTKVTCLSCELNPDRCQYYSVSFSQEA 464

Qy        189 ----LDSAGP-LPDYT---------------------------PPEIFSFESTTGFTL- 214
                  |  :|| || ||                          | :   |     |
Db        465 KYYQLRCSGPGLPLYTLHNSNNDKELRVLENNSDLDQVLQDVQMPSKKLDFIHLHGTKFW 524

Qy        215 YGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-I 270
              | |:  ||     |||| :| :| ||  |       |    ||||    |||   ::|
Db        525 YQMILPPH-FDKSKKYPLLLEVYAGPCSQ------KADAIFRLNWATYLASTENIIVASF 577

Qy        271 DNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSL 330
              | ||| ::| |   |   ::|  |::||:|  :   |:  |:|  |: | |||||||::
Db        578 DGRGSGYQGDKIMHAINRRLGTFEVEDQIEATRQF-SKMGFVDDKRIAIWGWSYGGYVTS 636

Qy        331 MALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              | |   | :|:  || |||: | :||: ||
Db        637 MVLGAGSGVFKCGIAVAPVSKWEYYDSVYT 666


RESULT 3
DPP4_HUMAN
ID   DPP4_HUMAN     STANDARD;      PRT;   766 AA.
AC   P27487;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE   antigen CD26) (TP103) (Adenosine deaminase complexing protein-2)
DE   (ADABP).
GN   DPP4 OR ADCP2 OR CD26.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92329551; PubMed=1352704;
RA   Misumi Y., Hayashi Y., Arakawa F., Ikehara Y.;
RT   "Molecular cloning and sequence analysis of human dipeptidyl
RT   peptidase IV, a serine proteinase on the cell surface.";
RL   Biochim. Biophys. Acta 1131:333-336(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=95012454; PubMed=7927537;
RA   Abbott C.A., Baker E., Sutherland G.R., McCaughan G.W.;
RT   "Genomic organization, exact localization, and tissue expression of
RT   the human CD26 (dipeptidyl peptidase IV) gene.";
RL   Immunogenetics 40:331-338(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Peripheral blood;
RX   MEDLINE=92325476; PubMed=1352530;
RA   Tanaka T., Camerini D., Seed B., Torimoto Y., Dang N.H., Kameoka J.,
RA   Dahlberg H.N., Schlossman S.F., Morimoto C.;
RT   "Cloning and functional expression of the T cell activation antigen
RT   CD26.";
RL   J. Immunol. 149:481-486(1992).
RN   [4]
RP   ERRATUM.
RX   MEDLINE=93171637; PubMed=8094732;
RA   Tanaka T.;
RL   J. Immunol. 150:2090-2090(1993).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Prostate;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [6]
RP   SEQUENCE OF 1-551 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=92165847; PubMed=1347043;
RA   Darmoul D., Lacasa M., Baricault L., Marguet D., Sapin C.,
RA   Trotot P., Barbat A.;
RT   "Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like
RT   colon cancer cell lines HT-29 and Caco-2. Cloning of the complete
RT   human coding sequence and changes of dipeptidyl peptidase IV mRNA
```

```
RT   levels during cell differentiation.";
RL   J. Biol. Chem. 267:4824-4833(1992).
RN   [7]
RP   SEQUENCE OF 545-766 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=91024044; PubMed=1977364;
RA   Darmoul D., Lacasa M., Chantret I., Swallow D., Trugnan G.;
RT   "Isolation of a cDNA probe for the human intestinal
RT   dipeptidylpeptidase IV and assignment of the gene locus DPP4 to
RT   chromosome 2.";
RL   Ann. Hum. Genet. 54:191-197(1990).
RN   [8]
RP   SEQUENCE OF 1-31 FROM N.A.
RX   MEDLINE=96067599; PubMed=7487939;
RA   Boehm S.K., Gum J.R. Jr., Erickson R.H., Hicks J.W., Kim Y.S.;
RT   "Human dipeptidyl peptidase IV gene promoter: tissue-specific
RT   regulation from a TATA-less GC-rich sequence characteristic of a
RT   housekeeping gene promoter.";
RL   Biochem. J. 311:835-843(1995).
RN   [9]
RP   PARTIAL SEQUENCE.
RC   TISSUE=Kidney;
RX   MEDLINE=93210468; PubMed=8096237;
RA   Morrison M.E., Vijayasaradhi S., Engelstein D., Albino A.P.,
RA   Houghton A.N.;
RT   "A marker for neoplastic progression of human melanocytes is a cell
RT   surface ectopeptidase.";
RL   J. Exp. Med. 177:1135-1143(1993).
CC   -!- FUNCTION: Removes N-terminal dipeptides sequentially from
CC       polypeptides having unsubstituted N-termini provided that the
CC       penultimate residue is proline. Plays a role in T cell activation.
CC   -!- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC       Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC       Xcc is neither Pro nor hydroxyproline.
CC   -!- SUBUNIT: Homodimer or heterodimer with Seprase (FAP).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC       a soluble form.
CC   -!- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC       by proteolytic processing.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD26 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd26.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U13735; AAB60646.1; -.
DR   EMBL; U13710; AAB60646.1; JOINED.
DR   EMBL; U13711; AAB60646.1; JOINED.
DR   EMBL; U13712; AAB60646.1; JOINED.
DR   EMBL; U13713; AAB60646.1; JOINED.
DR   EMBL; U13714; AAB60646.1; JOINED.
DR   EMBL; U13715; AAB60646.1; JOINED.
DR   EMBL; U13716; AAB60646.1; JOINED.
DR   EMBL; U13717; AAB60646.1; JOINED.
DR   EMBL; U13718; AAB60646.1; JOINED.
DR   EMBL; U13719; AAB60646.1; JOINED.
DR   EMBL; U13720; AAB60646.1; JOINED.
DR   EMBL; U13721; AAB60646.1; JOINED.
DR   EMBL; U13722; AAB60646.1; JOINED.
DR   EMBL; U13723; AAB60646.1; JOINED.
DR   EMBL; U13724; AAB60646.1; JOINED.
DR   EMBL; U13725; AAB60646.1; JOINED.
DR   EMBL; U13726; AAB60646.1; JOINED.
DR   EMBL; U13727; AAB60646.1; JOINED.
DR   EMBL; U13728; AAB60646.1; JOINED.
DR   EMBL; U13729; AAB60646.1; JOINED.
DR   EMBL; U13730; AAB60646.1; JOINED.
DR   EMBL; U13731; AAB60646.1; JOINED.
DR   EMBL; U13732; AAB60646.1; JOINED.
DR   EMBL; U13733; AAB60646.1; JOINED.
DR   EMBL; U13734; AAB60646.1; JOINED.
DR   EMBL; M74777; AAA51943.1; -.
DR   EMBL; BC013329; AAH13329.1; -.
DR   EMBL; M80536; AAA52308.1; -.
DR   EMBL; X60708; CAA43118.1; -.
DR   EMBL; S79876; AAB35614.1; -.
DR   PIR; S24313; CDHU26.
DR   PDB; 1N1M; 07-JAN-03.
DR   MEROPS; S09.003; -.
DR   Genew; HGNC:3009; DPP4.
DR   MIM; 102720; -.
DR   GO; GO:0004274; F:dipeptidyl-peptidase IV activity; TAS.
DR   GO; GO:0005209; F:plasma protein; TAS.
DR   GO; GO:0006955; P:immune response; TAS.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW   Transmembrane; Glycoprotein; Signal-anchor; 3D-structure.
FT   CHAIN         1    766       DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT                                (MDPP).
FT   CHAIN        39    766       DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT                                (SDPP).
FT   DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     28       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       29    766       EXTRACELLULAR (POTENTIAL).
FT   ACT_SITE    630    630       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    708    708       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    740    740       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     85     85       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     92     92       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    150    150       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    219    219       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    229    229       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    281    281       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    321    321       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    520    520       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    685    685       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      6      6       K -> R (IN REF. 6).
FT   CONFLICT      7      7       V -> I (IN REF. 1).
FT   CONFLICT    437    437       S -> I (IN REF. 1).
FT   CONFLICT    557    557       T -> I (IN REF. 7).
FT   CONFLICT    663    663       D -> E (IN REF. 7).

  Query Match           16.3%;  Score 315;  DB 1;  Length 766;
  Best Local Similarity  23.4%;  Pred. No. 2.9e-19;
  Matches  120;  Conservative   59;  Mismatches  146;  Indels  188;  Gaps   22;

Qy          1 EEDARSAGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH--- 56
              :||    |:  :| :|| |  ||  || |           : | : :::|| :|
Db        190 KEDIIYNGITDWVYEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSF 240

Qy         57 VTSPMLETRRADSFRYPKTGTANPKVTF-----------KMSEIMIDAEGRIIVDEVRR 104
               :  |:  :     ||| |  || | |               : | | |   :::
Db        241 YSDESLQYPKTVRVPYPKAGAVNPTVKFFVVNTDSLSSVTNATSIQITAPASMLIG---- 296

Qy        105 LVYFEGTKDSPLEHHLYVVSYVNPGEVT-----RLTDRGYSHSC----------CI--SQ 147
                        :|:|  |::     ::     |: :      |           |:   |
Db        297 -----------DHYLCDVTWATQERISLQWLRRIQNYSVMDICDYDESSGRWNCLVARQ 344

Qy        148 HCDF----FISKYSNQKNPHCV----SLYKLSSPED-------------DPTCKTKEFWA 186
              | :     :: ::     ||       | ||: | |:                | | ||  |
Db        345 HIEMSTTGWVGRF-RPSEPHFTLDGNSFYKIISNEEGYRHICYFQIDKKDCTFITKGTWE 403

Qy        187 TI---------------------------------------------------------- 188
```

```
Db        404 VIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYTKVTCLSCELNPERCQYYSVSPSK 463

Qy        189 ------LDSAGP-LPDYT----------------------------PPEIFSF--ESTTG 211
Db        464 EAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSALDKMLQNVQMPSKKLDFIILNETK 523

Qy        212 FTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVV 268
Db        524 F-WYQMILPPH-FDKSKKYPLLLDVYAGPCSQ------KADTVFRLNWATYLASTENIIV 575

Qy        269 V-IDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGY 327
Db        576 ASFDGRGSGYQGDKIMHAINRRLGTFEVEDQIEAARQF-SKMGFVDNKRIAIWGWSYGGY 634

Qy        328 LSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
Db        635 VTSMVLGSGSGVFKCGIAVAPVSRWEYYDSVYT 667
```

```
RESULT 4
YEA8_SCHPO
ID   YEA8_SCHPO     STANDARD;      PRT;   793 AA.
AC   O14073;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Putative dipeptidyl aminopeptidase C2E11.08 (EC 3.4.14.-).
GN   SPACUNK4.08 OR SPAC2E11.08.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like
CC       vacuoles (By similarity).
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL031180; CAA20138.1; -.
DR   PIR; T41703; T41703.
DR   GeneDB_SPombe; SPACUNK4.08; -.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hypothetical protein; Hydrolase; Aminopeptidase; Dipeptidase;
KW   Serine protease; Transmembrane; Glycoprotein; Signal-anchor.
FT   DOMAIN        1     24       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     25     45       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       46    793       LUMENAL (POTENTIAL).
FT   ACT_SITE    647    647       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    722    722       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    755    755       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD    101    101       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    136    136       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    246    246       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    299    299       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    303    303       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    324    324       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    336    336       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    377    377       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    384    384       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    407    407       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    535    535       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    761    761       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   793 AA;  91304 MW;  20B707F97F231463 CRC64;
```

```
  Query Match           16.2%;  Score 312.5;  DB 1;  Length 793;
  Best Local Similarity 22.9%;  Pred. No. 5e-19;
  Matches  114;  Conservative   62;  Mismatches  149;  Indels  173;  Gaps   17;

Qy          8 GVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV--------- 57
Db        214 GLTDWIYEEEVLSSPSTIWWSPDSD---------KIAFLKLNESEIPTYHYPLYTAELDP 264

Qy         58 TSPMLETRRADSFRYPKTGTANPKVTFKMSEI---------------------------- 89
Db        265 SLPEFDYNKDMAIKYPKPGNPNPSVSLFVADLNSNASSNFSLWHNEPLAEPVVQNVLWVN 324

Qy         90 ------------------MIDAEGRII------------------------------- 98
Db        325 TSSVLVQFTNRNSTCITARLLDTELKSIHTVKTECLEEGWYEVQQSAKMFPLNNSLVWEN 384

Qy         99 ----------VDEVRRL--------------------------------VYFEGTKDSP 115
Db        385 WSDGYFDILALDDYNHLAFIPFNGSSPIYLTSGAWDVTDGPIHIDGDFGNVYFLATLKDS 444

Qy        116 LEHHLYVVSYVNPGEVTRLTDRGYS---HSCCISQHCDFFISKYSNQKNPHCVSLYKLSS 172
Db        445 TERHLYYVS-LDTLEIYGITDNGEDEGYYSTSFSPFGDFYVLNYHGPDVP----WQELRS 499

Qy        173 PEDDPTCKTKEFWATILDSAGPLPDYTPPEI----FSFESTTGFTLYGMLYKPHDLQPGK 228
Db        500 TKDKDYCLSLETNSRLKQQ---LSSITLPSVEYGKLTFNDTT-FNF--MERRPRNFDVNK 553

Qy        229 KYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLG------YVVVVIDNRGSCHRGLKF 282
Db        554 KYPVLFFAYGGPGSQQV------AKLFRVDFQAYLASHPDFEFIVVTLDGRGTGFNGNAF 607

Qy        283 EGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRV 342
Db        608 RYSVSRHLGEWESYDQGQAGKFWAD-LPFVDENHVGIWGWSYGGYLTLKTL-ETQDVFSY 665

Qy        343 AIAGAPVTLWIFYDTGYT 360
```

```
                :| |||| |  ||: ||
Db        666 GMAVAPVTDWRLYDSVYT 683


RESULT 5
DPP4_MOUSE
ID   DPP4_MOUSE     STANDARD;      PRT;   760 AA.
AC   P28843;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE   antigen CD26) (Thymocyte-activating molecule) (THAM).
GN   DPP4 OR CD26.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Swiss; TISSUE=Thymus;
RX   MEDLINE=92129288; PubMed=1370813;
RA   Marguet D.A., Bernard A.-M., Vivier I., Darmoul D., Naquet P.,
RA   Pierres M.;
RT   "cDNA cloning for mouse thymocyte-activating molecule. A
RT   multifunctional ecto-dipeptidyl peptidase IV (CD26) included in a
RT   subgroup of serine proteases.";
RL   J. Biol. Chem. 267:2200-2208(1992).
RN   [2]
RP   REVISIONS.
RA   Marguet D.A.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B10.A; TISSUE=Liver;
RX   MEDLINE=95092780; PubMed=7999781;
RA   Bernard A.-M., Mattei M.-G., Pierres M., Marguet D.;
RT   "Structure of the mouse dipeptidyl peptidase IV (CD26) gene.";
RL   Biochemistry 33:15204-15214(1994).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SEQUENCE OF 1-20.
RX   MEDLINE=91302787; PubMed=1712807;
RA   Vivier I., Marguet D.A., Naquet P., Bonicel J., Black D., Li C.X.-Y.,
RA   Bernard A.-M., Gorvel J.-P., Pierres M.;
RT   "Evidence that thymocyte-activating molecule is mouse CD26
RT   (dipeptidyl peptidase IV).";
RL   J. Immunol. 147:447-454(1991).
CC   -!- FUNCTION: Removes N-terminal dipeptides sequentially from
CC       polypeptides having unsubstituted N-termini provided that the
CC       penultimate residue is proline.
CC   -!- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC       Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC       Xcc is neither Pro nor hydroxyproline.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC       a soluble form.
CC   -!- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC       by proteolytic processing.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X58384; CAA41274.1; -.
DR   EMBL; U12620; AAA82213.1; -.
DR   EMBL; U12599; AAA82213.1; JOINED.
DR   EMBL; U12600; AAA82213.1; JOINED.
DR   EMBL; U12601; AAA82213.1; JOINED.
DR   EMBL; U12602; AAA82213.1; JOINED.
DR   EMBL; U12603; AAA82213.1; JOINED.
DR   EMBL; U12604; AAA82213.1; JOINED.
DR   EMBL; U12605; AAA82213.1; JOINED.
DR   EMBL; U12606; AAA82213.1; JOINED.
DR   EMBL; U12607; AAA82213.1; JOINED.
DR   EMBL; U12608; AAA82213.1; JOINED.
DR   EMBL; U12609; AAA82213.1; JOINED.
DR   EMBL; U12610; AAA82213.1; JOINED.
DR   EMBL; U12611; AAA82213.1; JOINED.
DR   EMBL; U12612; AAA82213.1; JOINED.
DR   EMBL; U12613; AAA82213.1; JOINED.
DR   EMBL; U12614; AAA82213.1; JOINED.
DR   EMBL; U12615; AAA82213.1; JOINED.
DR   EMBL; U12616; AAA82213.1; JOINED.
DR   EMBL; U12617; AAA82213.1; JOINED.
DR   EMBL; U12618; AAA82213.1; JOINED.
DR   EMBL; U12619; AAA82213.1; JOINED.
DR   EMBL; BC022183; AAH22183.1; -.
DR   MEROPS; S09.003; -.
DR   MGD; MGI:94919; Dpp4.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW   Transmembrane; Glycoprotein; Signal-anchor.
FT   CHAIN         1    760       DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT                                (MDPP).
FT   CHAIN        37    760       DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT                                (SDPP).
FT   DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     28       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       29    760       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      295    466       CYS-RICH.
FT   ACT_SITE    624    624       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    702    702       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    734    734       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     83     83       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     90     90       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    113    113       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    213    213       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    223    223       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    315    315       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    328    328       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    514    514       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    679    679       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   760 AA;  87436 MW;  A5F644B46E4A3DF8 CRC64;

  Query Match           16.0%;  Score 309;  DB 1;  Length 760;
  Best Local Similarity  25.7%;  Pred. No. 9.5e-19;
  Matches  128;  Conservative   57;  Mismatches  154;  Indels  160;  Gaps   23;

Qy          1 EEDARSAGVATFVLQEE-FDRYSGYWWCPK------AETTPSGGKILRILYEENDESEVE 53
              ||:    |: :| :|| |  ||  || |       |:  :|  ::   :  :|||
Db        184 EENVIYNGITDWVYEEEVFGAYSALWWSPNNTFLAYAQFNDTGVPLIEYSF-YSDES--- 239

Qy         54 IIHVTSPMLETRRADSFRYPKTGTANPKVTF---------------------------- 84
                    |:  :     ||| |  || | |
Db        240 --------LQYPKTVWIPYPKAGAVNPTVKFFIVNIDSLSSSSSAAPIQIPAPASVARGD 291

Qy         85 -KMSEIMIDAEGRIIVDEVRRLVYFEG-----------TKDSPLE-HHLYV-----VSYV 126
               : :::   | || :  :||:  :             | : | |  |: :     |
Db        292 HYLCDVVWATEERISLQWLRRIQNYSVMAICDYDKINLTWNCPSEQQHVEMSTTGWVGRF 351

Qy        127 NPGEVTRLTD-----------RGYSHSCCI---SQHCDF----------FISK 155
              | |    :|           || | |      : | |                ::
Db        352 RPAEPHFTSDGSSFYKIISDKDGYKHICHFPKDKKDCTFITKGAWEVISIEALTSDYLYY 411

Qy        156 YSNQ--KNPHCVSLYK-----------LSSPEDDPTCK------TKEFWATILDSAGP-L 195
              |||  : |   :|||           ||   :   |:      :||      |   || |
Db        412 ISNQYKEMPGGRNLYKIQLTDHTNVKCLSCDLNPERCQYYAVSFSKEAKYYQLGCWGPGL 471

Qy        196 PDYT----------------------------PPEIFSF--ESTTGFTLYGMLYKPHDLQ 225
              | ||                          | :   |   : | |  | |:  ||
Db        472 PLYTLHRSTDHKELRVLEDNSALDRMLQDVQMPSKKLDFIVLNETRF-WYQMILPPH-FD 529

Qy        226 PGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLK 281
              |||| :| :| || |    |    ||||   |||   ::|   | ||| ::| |
Db        530 KSKKYPLLLDVYAGPCSQ------KADASFRLNWATYLASTENIIVASFDGRGSGYQGDK 583

Qy        282 FEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFR 341
                |   ::| :|::||:|  :    :  |:|  || | |||||||:: | |   | :|:
Db        584 IMHAINRRLGTLEVEDQIEAARQFV-KMGFVDSKRVAIWGWSYGGYVTSMVLGSGSGVFK 642

Qy        342 VAIAGAPVTLWIFYDTGYT 360
              || |||: | :||: ||
Db        643 CGIAVAPVSRWEYYDSVYT 661


RESULT 6
DPP4_RAT
ID   DPP4_RAT       STANDARD;      PRT;   767 AA.
AC   P14740;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE   antigen CD26) (GP110 glycoprotein) (Bile canaliculus domain-specific
DE   membrane glycoprotein).
GN   DPP4 OR CD26.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89123496; PubMed=2563382;
RA   Ogata S., Misumi Y., Ikehara Y.;
RT   "Primary structure of rat liver dipeptidyl peptidase IV deduced from
RT   its cDNA and identification of the NH2-terminal signal sequence as
RT   the membrane-anchoring domain.";
RL   J. Biol. Chem. 264:3596-3601(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88068516; PubMed=3479775;
RA   Hong W., Doyle D.;
RT   "cDNA cloning for a bile canaliculus domain-specific membrane
RT   glycoprotein of rat hepatocytes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:7962-7966(1987).
RN   [3]
RP   SIGNAL-ANCHOR.
RX   MEDLINE=90338089; PubMed=1974258;
RA   Hong W., Doyle D.;
RT   "Molecular dissection of the NH2-terminal signal/anchor sequence of
RT   rat dipeptidyl peptidase IV.";
RL   J. Cell Biol. 111:323-328(1990).
RN   [4]
RP   SEQUENCE OF 281-302.
RC   TISSUE=Kidney;
RX   MEDLINE=94128239; PubMed=7905271;
RA   Iwaki-Egawa S., Watanabe Y., Fujimoto Y.;
RT   "N-terminal amino acid sequence of the 60-kDa protein of rat kidney
RT   dipeptidyl peptidase IV.";
RL   Biol. Chem. Hoppe-Seyler 374:973-975(1993).
CC   -!- FUNCTION: REMOVES N-TERMINAL DIPEPTIDES SEQUENTIALLY FROM
CC       POLYPEPTIDES HAVING UNSUBSTITUTED N-TERMINI PROVIDED THAT THE
CC       PENULTIMATE RESIDUE IS PROLINE.
CC   -!- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC       Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC       Xcc is neither Pro nor hydroxyproline.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS IN
CC       A SOLUBLE FORM.
CC   -!- PTM: THE SOLUBLE FORM (SDPP) DERIVES FROM THE MEMBRANE FORM (MDPP)
CC       BY PROTEOLYTIC PROCESSING.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04591; AAA41096.1; -.
DR   EMBL; J02997; AAA41272.1; -.
DR   PIR; A39914; A39914.
DR   MEROPS; S09.003; -.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW   Transmembrane; Glycoprotein; Signal-anchor.
FT   CHAIN         1    767       DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT                                (MDPP).
FT   CHAIN        37    767       DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT                                (SDPP).
FT   CHAIN       281    767       DIPEPTIDYL PEPTIDASE IV 60 KDA SOLUBLE
FT                                FORM.
FT   DOMAIN       1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     7     28       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN      29    767       EXTRACELLULAR (POTENTIAL).
FT   ACT_SITE   631    631       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   709    709       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   741    741       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD    83     83       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    90     90       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   148    148       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   217    217       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   227    227       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   319    319       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   521    521       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   686    686       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    38     38       A -> R (IN REF. 2).
FT   CONFLICT   183    183       I -> T (IN REF. 2).
```

```
FT   CONFLICT    332    332       T -> N (IN REF. 2).
FT   CONFLICT    352    352       C -> V (IN REF. 2).
FT   CONFLICT    394    394       V -> D (IN REF. 2).
FT   CONFLICT    562    562       L -> F (IN REF. 2).
FT   CONFLICT    624    624       R -> Q (IN REF. 2).
SQ   SEQUENCE   767 AA;  88003 MW;  AA17C0EC6F0E4652 CRC64;

  Query Match             15.4%;  Score 297.5;  DB 1;  Length 767;
  Best Local Similarity   24.5%;  Pred. No. 9.4e-18;
  Matches  121;  Conservative   60;  Mismatches  152;  Indels  161;  Gaps   22;

Qy          8 GVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLETRR 66
              |:  :| :|| |  ||  ||      :|:|  :    : :    :|    :   |:  :
Db        195 GINDWVYEEEIFGAYSALWW------SPNGTFLAYAQFNDTGVPLIEYSFYSDESLQYPK 248

Qy         67 ADSFRYPKTGTANPKVTF----------------------------KMSEIMIDAEGR 96
                ||| |  || | |                            : ::   :| |
Db        249 TVWIPYPKAGAVNPTVKFFIVNTDSLSSTTTTIPMQITAPASVTTGDHYLCDVAWVSEDR 308

Qy         97 IIVDEVRRLVYF--------------------EGTKDS---------PLEHHLYVVSYV 126
              |  :  :||:  :                  |  : |         | | |     :
Db        309 ISLQWLRRIQNYSVMAICDYDKTTLVWNCPTTQEHIETSATGWCGRFRPAEPH-----FT 363

Qy        127 NPGE--VTRLTDR-GYSHSCCI------SQHCDF-----------------FISKYSNQ- 159
              : |       ::|: || | |          | | |                ::   ||:
Db        364 SDGSSFYKIVSDKDGYKHICQFQKDRKPEQVCTFITKGAWEVISIEALTSDYLYYISNEY 423

Qy        160 -KNPHCVSLYK-----------LSSPEDDPTCK------TKEFWATILDSAGP-LPDYT- 199
               : |   :|||           ||   :   |:      :||      |   || || ||
Db        424 KEMPGGRNLYKIQLTDHTNKKCLSCDLNPERCQYYSVSLSKEAKYYQLGCRGPGLPLYTL 483

Qy        200 -----PPEIFSFESTTGFT----------------------LYGMLYKPHDLQPGKKYP 231
                  |:   |  :                          | |:  ||     ||||
Db        484 HRSTDQKELRVLEDNSALDKMLQDVQMPSKKLDFIVLNETRFWYQMILPPH-FDKSKKYP 542

Qy        232 TVLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLKFEGAFK 287
               :: :| ||  |    |   ||||    |||   ::|   | ||| ::| |   |
Db        543 LLIDVYAGPCSQ------KADAAFRLNWATYLASTENIIVASFDGRGSGYQGDKIMHAIN 596

Qy        288 YKMGQIEIDDQVEGL-QYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAG 346
               ::| :|::||:|   |:|  :  |:|  || | |||||||:: | |   | :|:  ||
Db        597 KRLGTLEVEDQIEAARQFL--KMGFVDSKRVAIWGWSYGGYVTSMVLGSGSGVFKCGIAV 654

Qy        347 APVTLWIFYDTGYT 360
              |||: | :||: ||
Db        655 APVSRWEYYDSVYT 668


RESULT 7
DAP2_YEAST
ID   DAP2_YEAST     STANDARD;      PRT;   818 AA.
AC   P18962;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Dipeptidyl aminopeptidase B (EC 3.4.14.-) (DPAP B) (YSCV).
GN   DAP2 OR YHR028C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89174971; PubMed=2647766;
RA   Roberts C.J., Pohlig G., Rothman J.H., Stevens T.H.;
RT   "Structure, biosynthesis, and localization of dipeptidyl
RT   aminopeptidase B, an integral membrane glycoprotein of the yeast
RT   vacuole.";
RL   J. Cell Biol. 108:1363-1373(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=94378003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
CC       VACUOLES.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X15484; CAA33512.1; -.
DR   EMBL; U10399; AAB68879.1; -.
DR   PIR; S46780; A30107.
DR   MEROPS; S09.006; -.
DR   SGD; S0001070; DAP2.
DR   GO; GO:0000329; C:vacuolar membrane (sensu Fungi); IDA.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW   Transmembrane; Glycoprotein; Signal-anchor.
FT   DOMAIN        1     29       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     30     45       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       46    818       LUMENAL (POTENTIAL).
FT   ACT_SITE    679    679       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    756    756       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    789    789       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     63     63       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     79     79       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    139    139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    372    372       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    392    392       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    421    421       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    738    738       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     83     83       Q -> H (IN REF. 1).
FT   CONFLICT    125    125       S -> N (IN REF. 1).
FT   CONFLICT    182    188       FEEIGNE -> LRRLET (IN REF. 1).
FT   CONFLICT    200    200       D -> N (IN REF. 1).
FT   CONFLICT    366    375       TSNVVRNESS -> DFKRGKERKF (IN REF. 1).
FT   CONFLICT    808    818       AKRAFDGQFVK -> QSVLSMGNLTNELTIYSSSHRDIHKT
FT                                FSYLHTMYI (IN REF. 1).
SQ   SEQUENCE   818 AA;  93404 MW;  318F450445375BD3 CRC64;

  Query Match             14.5%;  Score 280.5;  DB 1;  Length 818;
  Best Local Similarity   22.0%;  Pred. No. 3e-16;
  Matches  114;  Conservative   50;  Mismatches  155;  Indels  199;  Gaps   19;

Qy          8 GVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEV-EIIHVTSPMLETR 65
              |   :| :|| |:     || |  :          : : : ||||| | |     |
Db        234 GKPDWVYEEEVFEDDKAAWWSPTGD---------YLAFLKIDESEVGEFI---IPYYVQD 281

Qy         66 RAD------SFRYPKTGTANP------------------------------------ 80
                 |      | :|||:|| ||
Db        282 EKDIYPEMRSIKYPKSGTPNPHAELWVYSMKDGTSFHPRISGNKKDGSLLITEVTWVGNG 341
```

```
Qy          81 ----KVTFKMSEI----MID------------------------------------- 92
               |  | :  |:|    :||
Db         342 NVLVKTTDRSSDILTVFLIDTIAKTSNVVRNESSNGGWWEITHNTLFIPANETFDRPHNG 401

Qy          93 ---------------------------------AEGRIIVDEVRRLVYFEGTKDSPLEH 118
                                                |  :  |  :   :||  |: |  |
Db         402 YVDILPIGGYNHLAYFENSNSSHYKTLTEGKWEVVNGPLAFDSMENRLYFISTRKSSTER 461

Qy         119 HLYVVSYVNPGEVTRLTDRGYS--HSCCISQHCDFFISKYSNQKNPH------------ 163
               |:| :   :| |:  :||      :    |    | :  |   | |:
Db         462 HVYYIDLRSPNEIIEVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEK 521

Qy         164 C-------VSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYG 216
               |       ||| |   |                :|     | || |     |  |   :|
Db         522 CDKGNVLGKSLYHLEKNE------------VLTKI--LEDYAVPRKSFRELNLGKDEFG 566

Qy         217 --MLYKPHDLQP-------GKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN----TLASL 263
                 :|   ::: |            ||   | ||||  | |      || | :      : |
Db         567 KDILVNSYEILPNDFDETLSDHYPVFFFAYGGPNSQQV------VKTFSVGFNEVVASQL 620

Qy         264 GYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWS 323
               :|||:| ||:  :|  |    : ::|  |  ||:       |   |:|  :: : |||
Db         621 NAIVVVVDGRGTGFKGQDFRSLVRDRLGDYEARDQISAASLYGS-LTFVDPQKISLFGWS 679

Qy         324 YGGYLSLMALMQRSD-IFRVAIAGAPVTLWIFYDTGYT 360
               |||||:|  | :     |:  :: |||| | |||: ||
Db         680 YGGYLTLKTLEKDGGRHFKYGMSVAPVTDWRFYDSVYT 717
```

RESULT 8
SEPR_MOUSE
ID SEPR_MOUSE STANDARD; PRT; 761 AA.
AC P97321;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Seprase (EC 3.4.21.-) (Fibroblast activation protein alpha) (Integral
DE membrane serine protease).
GN FAP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC STRAIN=BALB/c; TISSUE=Embryo;
RX MEDLINE=97284459; PubMed=9139873;
RA Niedermeyer J., Scanlan M.J., Garin-Chesa P., Daiber C., Fiebig H.H.,
RA Old L.J., Rettig W.J., Schnapp A.;
RT "Mouse fibroblast activation protein: molecular cloning, alternative
RT splicing and expression in the reactive stroma of epithelial
RT cancer.";
RL Int. J. Cancer 71:383-389(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Breast;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: May have a role in tissue remodeling during development
CC and wound healing, and contribute to invasiveness in malignant
CC cancers.
CC -!- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and
CC type IV collagen, but not native type I or type IV collagen. Does
CC not cleave laminin, fibronectin, fibrin or casein.
CC -!- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is
CC inactive (By similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=P97321-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P97321-2; Sequence=VSP_005368;
CC Name=3;
CC IsoId=P97321-3; Sequence=VSP_005369;
CC -!- TISSUE SPECIFICITY: Detected in fibroblasts, in placenta, uterus,
CC embryos from day 7-19 and in new-born mice (P1).
CC -!- SIMILARITY: Belongs to peptidase family S9B.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; Y10007; CAA71116.1; -.
DR EMBL; BC019190; AAH19190.1; -.
DR MEROPS; S09.007; -.
DR MGD; MGI:109608; Fap.
DR InterPro; IPR002469; DPPIV_N_term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002471; Prol_endopep_ser.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00930; DPPIV_N_term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane; Signal-anchor;
KW Glycoprotein; Alternative splicing.
FT DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 26 761 EXTRACELLULAR (POTENTIAL).
FT ACT_SITE 624 624 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 702 702 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 734 734 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 31 35 Missing (in isoform 2).
FT /FTId=VSP_005368.
FT VARSPLIC 31 63 Missing (in isoform 3).
FT /FTId=VSP_005369.
FT CONFLICT 737 737 S -> L (IN REF. 2).
SQ SEQUENCE 761 AA; 87944 MW; 9174C3AEDA213B25 CRC64;

Query Match 13.4%; Score 259; DB 1; Length 761;
Best Local Similarity 21.2%; Pred. No. 2e-14;
Matches 106; Conservative 66; Mismatches 149; Indels 178; Gaps 18;

```
Qy           8 GVATFVLQEEF--DRYSGYWWCPKAETTPSGGKILRILYEENDESEVETIHVT------S 59
               |:  :| :||    :|:  || |         || |   | | ::|:: ||  :
Db         195 GIPDWVYEEEMLATKYA-LWWSP-------DGKFL--AYVEFNDSDIPIIAYSYYGDGQY 244
```

```
Qy          60 PMLETRRADSFRYPKTGTANPKV----------------------------------TFK 85
               |      |  :  ||| |  || |                                   |:
Db         245 P-----RTINIPYPKAGAKNPVVRVFIVDTTYPHHVGPMEVPVPEMIASSDYYFSWLTWV 299

Qy          86 MSE------------------------------------------------------------ 88
               ||
Db         300 SSERVCLQWLKRVQNVSVLSICDFREDWHAWECPKNQEHVEESRTGWAGGFFVSTPAFSQ 359

Qy          89 -------IMIDAEG----RIIVDEVRRLV-----------YFEGTKDS-----------P 115
                    |  | :|       | | |   :             |  |:||           |
Db         360 DATSYYKIFSDKDGYKHIHYIKDTVENAIQITSGKWEAIYIFRVTQDSLFYSSNEFEGYP 419

Qy         116 LEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQH-----CDFFISKYSNQKNPHCVSLY-- 168
                ::| :|  |      :          |:: |     | :: : :| :   :  :  |
Db         420 GRRNIYRISIGNSPPSKK----------CVTCHLRKERCQYYTASFSYKAKYYALVCYGP 469

Qy         169 --KLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEI-FSFESTTGFTLYGMLYKPHDLQ 225
                  :|:  |  |  :  :       :    | :   |::         | |  :  :  |
Db         470 GLPISTLHDGRTDQEIQVLEENKELENSLRNIQLPKVEIKKLKDGGLTFWYKMILPPQFD 529

Qy         226 PGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASL----GYVVVVIDNRGSCHRGLK 281
                ||||  :: :||||  | | :        | :| :  |    | |: ::| ||:  :| |
Db         530 RSKKYPLLIQVYGGPCSQSVKS------VFAVNWITYLASKEGIVIALVDGRGTAFQGDK 583

Qy         282 FEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFR 341
               |  |    |:|  |::||:  ::       ||| :|: | ||||||:| :||   : :|:
Db         584 FLHAVYRKLGVYEVEDQLTAVRKFI-EMGFIDEERIAIWGWSYGGYVSSLALASGTGLFK 642

Qy         342 VAIAGAPVTLWIFYDTGYT 360
                  || |||: | :| : |:
Db         643 CGIAVAPVSSWEYYASIYS 661
```

```
RESULT 9
SEPR_HUMAN
ID   SEPR_HUMAN     STANDARD;      PRT;   760 AA.
AC   Q12884; O00199; Q99998; Q9UID4;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Seprase (EC 3.4.21.-) (Fibroblast activation protein alpha) (Integral
DE   membrane serine protease) (170-kDa melanoma membrane-bound
DE   gelatinase).
GN   FAP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Fibroblast;
RX   MEDLINE=94261645; PubMed=7911242;
RA   Scanlan M.J., Raj B.K.M., Calvo B., Garin-Chesa P., Sanz-Moncasi M.P.,
RA   Healey J.H., Old L.J., Rettig W.J.;
RT   "Molecular cloning of fibroblast activation protein alpha, a member of
RT   the serine protease family selectively expressed in stromal
RT   fibroblasts of epithelial cancers.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:5657-5661(1994).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Melanoma;
RX   MEDLINE=97388251; PubMed=9247085;
RA   Goldstein L.A., Ghersi G., Pineiro-Sanchez M.L., Salamone M., Yeh Y.,
RA   Flessate D., Chen W.-T.;
RT   "Molecular cloning of seprase: a serine integral membrane protease
RT   from human melanoma.";
RL   Biochim. Biophys. Acta 1361:11-19(1997).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND SEQUENCE OF 220-229; 461-472 AND
RP   511-518.
RC   TISSUE=Melanoma;
RX   MEDLINE=97218181; PubMed=9065413;
RA   Pineiro-Sanchez M.L., Goldstein L.A., Dodt J., Howard L., Yeh Y.,
RA   Chen W.-T.;
RT   "Identification of the 170-kDa melanoma membrane-bound gelatinase
RT   (seprase) as a serine integral membrane protease.";
RL   J. Biol. Chem. 272:7595-7601(1997).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Melanoma;
RX   MEDLINE=20112818; PubMed=10644713;
RA   Goldstein L.A., Chen W.-T.;
RT   "Identification of an alternatively spliced seprase mRNA that encodes
RT   a novel intracellular isoform.";
RL   J. Biol. Chem. 275:2554-2559(2000).
RN   [5]
RP   SEQUENCE OF 192-208; 220-240 AND 510-521.
RX   MEDLINE=94327249; PubMed=7519584;
RA   Rettig W.J., Su S.L., Fortunato S.R., Scanlan M.J., Raj B.K.M.,
RA   Garin-Chesa P., Healey J.H., Old L.J.;
RT   "Fibroblast activation protein: purification, epitope mapping and
RT   induction by growth factors.";
RL   Int. J. Cancer 58:385-392(1994).
CC   -!- FUNCTION: May have a role in tissue remodeling during development
CC       and wound healing, and may contribute to invasiveness in malignant
CC       cancers.
CC   -!- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and
CC       type IV collagen, but not native type I or type IV collagen. Does
CC       not cleave laminin, fibronectin, fibrin or casein.
CC   -!- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is
CC       inactive.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Found in cell
CC       surface lamellipodia, invadopodia and on shed vesicles.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1; Synonyms=L;
CC         IsoId=Q12884-1; Sequence=Displayed;
CC         Note=Major isoform;
CC       Name=2; Synonyms=S, Truncated;
CC         IsoId=Q12884-2; Sequence=VSP_005367;
CC   -!- TISSUE SPECIFICITY: Fibroblast-specific.
CC   -!- INDUCTION: In fibroblasts at times and sites of tissue remodeling
CC       during development, tissue repair, and carcinogenesis.
CC   -!- PTM: N-glycosylated.
CC   -!- PTM: The N-terminus may be blocked.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U09278; AAB49652.1; -.
DR   EMBL; U76833; AAC51668.1; -.
DR   EMBL; AF007822; AAF21600.1; -.
DR   MEROPS; S09.007; -.
DR   Genew; HGNC:3590; FAP.
DR   MIM; 600403; -.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Protease; Serine protease; Transmembrane; Signal-anchor;
KW   Glycoprotein; Alternative splicing.
FT   DOMAIN        1      4       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      5     25       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       26    760       EXTRACELLULAR (POTENTIAL).
FT   ACT_SITE    624    624       CHARGE RELAY SYSTEM (BY SIMILARITY).
```

```
FT   ACT_SITE    702    702       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    734    734       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     49     49       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     92     92       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     99     99       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    314    314       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    679    679       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC      1    521       Missing (in isoform 2).
FT                                /FTId=VSP_005367.
FT   CONFLICT    207    207       P -> A (IN REF. 2).
FT   CONFLICT    229    229       K -> T (IN REF. 2).
FT   CONFLICT    354    354       R -> T (IN REF. 2).
SQ   SEQUENCE   760 AA;  87820 MW;  A0D34B4801BE07EA CRC64;

  Query Match             13.3%;  Score 257;  DB 1;  Length 760;
  Best Local Similarity   22.7%;  Pred. No. 2.9e-14;
  Matches  110;  Conservative   68;  Mismatches  158;  Indels  148;  Gaps   20;

Qy          8 GVATFVLQEEF--DRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLET- 64
              |:  :| :||    :|:  || |         || |   | | :: :: :|  :    |
Db        195 GIPDWVYEEEMLPTKYA-LWWSP-------NGKFL--AYAEFNDKDIPVIAYSYYGDEQY 244

Qy         65 RRADSFRYPKTGTANPKVTFKM----------SEIMIDA----------------EGRI 97
               |  :  ||| |  || |   :           |: : |                : |:
Db        245 PRTINIPYPKAGAKNPVVRIFIIDTTYPAYVGPQEVPVPAMIASSDYYFSWLTWVTDERV 304

Qy         98 IVDEVRRLVYFE----------GTKDSP-LEHHLY--------------VVSYVNPGE 130
               :  ::|:                | | |  : |:                | ||
Db        305 CLQWLKRVQNVSVLSICDFREDWQTWDCPKTQEHIEESRTGWAGGFFVSRPVFSYDAISY 364

Qy        131 VTRLTDR-GYSHSCCI----------------------SQHCDFFIS----KYSNQKNP 162
                 :|: || |   |                      :|   |: |    :|  ::|
Db        365 YKIFSDKDGYKHIHYIKDTVENAIQITSGKWEAINIFRVTQDSLFYSSNEFEEYPGRRNI 424

Qy        163 HCVSLYKLSSPEDDPTCKTKE-----FWATILDSA--------GP-LPDYT--------- 199
              : :|:      :   ||  ::     : |:  | |         || :|  |
Db        425 YRISIGSYPPSKKCVTCHLRKERCQYYTASFSDYAKYYALVCYGPGIPISTLHDGRTDQE 484

Qy        200 -------------------PPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGP 240
                              | |          ||:  :  |      |||| :: :||||
Db        485 IKILEENKELENALKNIQLPKEEIKKLEVDEITLWYKMILPPQFDRSKKYPLLIQVYGGP 544

Qy        241 QVQLVNNRFKGVKYFRLNTLASL----GYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEID 296
              | | :       | :|  :: |     | |: ::| ||:  :| |    |    |:|  |::
Db        545 CSQSVRS------VFAVNWISYLASKEGMVIALVDGRGTAFQGDKLLYAVYRKLGVYEVE 598

Qy        297 DQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYD 356
              ||:  ::       |||  |: | |||||||:| :||   : :|:  || |||: | :|
Db        599 DQITAVRKFI-EMGFIDEKRIAIWGWSYGGYVSSLALASGTGLFKCGIAVAPVSSWEYYA 657

Qy        357 TGYT 360
              : ||
Db        658 SVYT 661


RESULT 10
ST13_YEAST
ID   ST13_YEAST     STANDARD;      PRT;   931 AA.
AC   P33894;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Dipeptidyl aminopeptidase A (EC 3.4.14.-) (DPAP A) (YSCIV).
GN   STE13 OR YCI1 OR YOR219C OR YOR50-9.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95066382; PubMed=7975897;
RA   Anna-Arriola S.S., Herskowitz I.;
RT   "Isolation and DNA sequence of the STE13 gene encoding dipeptidyl
RT   aminopeptidase.";
RL   Yeast 10:801-810(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB320;
RA   Flanagan C.A., Thorner J.;
RT   "STE13.";
RL   (In) Gething M.-J., Novick P., Stevens T.H., Rothblatt J. (eds.);
RL   Guidebook to the yeast secretory pathway, pp.1-1, Oxford University
RL   Press, Oxford (1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1679;
RX   MEDLINE=96437977; PubMed=8840505;
RA   Galisson F., Dujon B.;
RT   "Sequence and analysis of a 33 kb fragment from the right arm of
RT   chromosome XV of the yeast Saccharomyces cerevisiae.";
RL   Yeast 12:877-885(1996).
CC   -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC MATURATION OF THE
CC       ALPHA-FACTOR PRECURSOR.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
CC       VACUOLES.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L21944; AAA35119.1; -.
DR   EMBL; U08230; AAA17897.1; -.
DR   EMBL; X92441; CAA63182.1; -.
DR   EMBL; Z75127; CAA99437.1; -.
DR   PIR; A49737; A49737.
DR   MEROPS; S09.005; -.
DR   SGD; S0005745; STE13.
DR   GO; GO:0005802; C:Golgi trans-face; IDA.
DR   GO; GO:0004177; F:aminopeptidase activity; IDA.
DR   GO; GO:0007323; P:peptide pheromone maturation; IDA.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; FALSE_NEG.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW   Transmembrane; Glycoprotein; Signal-anchor; Pheromone response.
FT   DOMAIN        1    119       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    120    140       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN      141    931       LUMENAL (POTENTIAL).
FT   ACT_SITE    785    785       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    863    863       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    896    896       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD    377    377       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    814    814       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   931 AA;  107200 MW;  81AF70094093C023 CRC64;

  Query Match             12.9%;  Score 250;  DB 1;  Length 931;
  Best Local Similarity   24.3%;  Pred. No. 1.5e-13;
  Matches  101;  Conservative   58;  Mismatches  162;  Indels   94;  Gaps   15;

Qy         15 QEEFDRYSGYWWCP---KAETTPSGGKILRILYEENDESEVEIIHVTS------------ 59
              |::   |:| |  |   : | |    ||| :   :   |::  :  |:
Db        434 QKDSILYNGKWISPDTFRFEITDRNSKILDVKVYDIPSSQMLTVRNTNSNLFNGWIEKTK 493

Qy         60 ------PMLETRR---------ADS------FRYPKTGTANPKVTFKMSEIMIDAEGRII 98
                    | | :|         |||      | || |  |   :       :   | :
Db        494 DILSIPPKPELKRMDYGYIDIHADSRGFSHLFYYP-TVFAKEPIQLTKGNWEVTGNGIVG 552
```

```
Qy         99 VDEVRRLVYFEGTKDSPLEHHLYVVSYV-----------NPGEVTRLTDRGYSHSCCIS 146
Db        553 YEYETDTIFFTANEIGVMSQHLYSISLTDSTTQNTFQSLQNP------SDKYDFYDFELS 606

Qy        147 QHCDFFISKYSNQKNP---------------HCVSLYKLSSPEDDPTCKTKEFWATILD 190
Db        607 SSARYAISKKLGPDTPIKVAGPLTRVLNVAEIHDDSILQLTKDE-----KFKE------- 654

Qy        191 SAGPLPDYTPPEIFSFESTT---GFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
Db        655 ---KIKNYDLP-ITSYKTMVLDDGVEINYIEIKPANLNPKKKYPILVNIYGGPGSQTFTT 710

Qy        248 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVE-GLQYLA 306
Db        711 --KSSLAFEQAVVSGLDVIVLQIEPRGTGGKGWSFRSWAREKLGYWEPRDITEVTKKFIQ 768

Qy        307 SRYDFIDLDRVGIHGWSYGGYLSLMAL-MQRSDIFRVAIAGAPVTLWIFYDTGYT 360
Db        769 RNSQHIDESKIAIWGWSYGGFTSLKTVELDNGDTFKYAMAVAPVTNWTLYDSVYT 823
```

RESULT 11
DPP6_BOVIN
ID DPP6_BOVIN STANDARD; PRT; 863 AA.
AC P42659;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
DE (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
DE (Dipeptidyl aminopeptidase-related protein) (DPPX).
GN DPP6.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S), AND PARTIAL SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=92108018; PubMed=1729689;
RA Wada K., Yokotani N., Hunter C., Doi K., Wenthold R.J., Shimasaki S.;
RT "Differential expression of two distinct forms of mRNA encoding
RT members of a dipeptidyl aminopeptidase family.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
CC -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
CC FUNCTION. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=DPPX-L;
CC IsoId=P42659-1; Sequence=Displayed;
CC Name=DPPX-S;
CC IsoId=P42659-2; Sequence=VSP_005364;
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN. DPPX-L
CC IS EXPRESSED EXCLUSIVELY IN THE BRAIN WHEREAS DPPX-S IS FOUND IN
CC BRAIN, KIDNEY, OVARY AND TESTIS.
CC -!- SIMILARITY: Belongs to peptidase family S9B.
CC ---

CC ---
DR EMBL; M76428; AAC41622.1; -.
DR EMBL; M76429; AAC41623.1; -.
DR PIR; A41793; A41793.
DR MEROPS; S09.973; -.
DR InterPro; IPR002469; DPPIV_N_term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00930; DPPIV_N_term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
KW Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
FT DOMAIN 1 93 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 94 114 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 115 863 EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 533 533 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 811 811 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 79 MASLYQRFTGKINTSRSFPAPPEASRLLGGQGPEEDGAGPK
FT PLGAQAPAAAPRERGGGGGAGGRPRFQYQARSDCDDED ->
FT MTTAKEPNASGKSVQQQEQ (in isoform DPPX-S).
FT /FTId=VSP_005364.
SQ SEQUENCE 863 AA; 96556 MW; 23DBA792B841A39D CRC64;

Query Match 11.4%; Score 221; DB 1; Length 863;
Best Local Similarity 21.0%; Pred. No. 4.5e-11;
Matches 103; Conservative 58; Mismatches 169; Indels 160; Gaps 16;

```
Qy          8 GVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV---TSPMLE 63
Db        285 GLSDWLYEEEILKTHIAHWWSPDG---------TRLAYATINDSRVPVMELPTYTGSVYP 335

Qy         64 TRRADSFRYPKTGTANPKVTF--------------------KMSEIMI------------ 91
Db        336 T--AKPYHYPKAGCENPSISLHVIGLNGPTHDLEMTPPDDPRMREYYITMVKWATSTKVA 393

Qy         92 ----------------DA----------------------------EGR----------- 96
Db        394 VNWLSRAQNVSILTLCDATTGVCTKKHEDESEAWLHRQNEEPVFSKDGRKFFFVRAIPQG 453

Qy         97 ----------------------------------IIVDEVRRLVYFEGTKDSPLEHHLY 121
Db        454 GQGKFYHITVSSSQPNSSNDNIQSITSGDWDVTKILSYDEKRSQIYFLSTEDLPRRRQLY 513

Qy        122 VVSYVNPGEVTR------LTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPED 175
Db        514 SASTV--GSFNRQCLSCDLVDNCTYFSASFSPGADFFLLKCEGPGVP-TVSVHNTTDKKK 570

Qy        176 DPTCKTKE-FWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVL 234
Db        571 MFDLETNEHVQKAISDRQMPKVEYRKIE------TDDYNLPIQILKPATFTDTAHYPLLL 624

Qy        235 FIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIE 294
Db        625 VVDGTPGSQSVAEKF--AVTWETVMVSSHGAVVVKCDGRGSGFQGTRLLHEVRRRLGSLE 682

Qy        295 IDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSD----IFRVAIAGAPVT 350
Db        683 EKDQMEAVRVML-KEPYIDKTRVAVFGKDYGGYLSTYLLPAKGDGQAPVFSCGSALSPIT 741

Qy        351 LWIFYDTGYT 360
Db        742 DFKLYASAFS 751
```

RESULT 12
DPP6_RAT
ID DPP6_RAT STANDARD; PRT; 859 AA.
AC P46101;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
DE (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
DE (Dipeptidyl aminopeptidase-related protein) (DPPX).

```
GN   DPP6.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S).
RC   TISSUE=Brain;
RX   MEDLINE=92108018; PubMed=1729689;
RA   Wada K., Yokotani N., Hunter C., Doi K., Wenthold R.J., Shimasaki S.;
RT   "Differential expression of two distinct forms of mRNA encoding
RT   members of a dipeptidyl aminopeptidase family.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
CC   -!- FUNCTION: May be involved in the physiological processes of brain
CC       function. Has no dipeptidyl aminopeptidase activity. The lack of
CC       activity may be due to the substitution of an aspartate residue
CC       for the serine residue in the proposed catalytic triad.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=DPPX-L;
CC         IsoId=P46101-1; Sequence=Displayed;
CC       Name=DPPX-S;
CC         IsoId=P46101-2; Sequence=VSP_005366;
CC   -!- TISSUE SPECIFICITY: DPPX-S IS EXPRESSED IN BRAIN AND SOME
CC       PERIPHERAL TISSUES INCLUDING KIDNEY, OVARY, AND TESTIS; IN
CC       CONTRAST DPPX-L IS EXPRESSED ALMOST EXCLUSIVELY IN BRAIN.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M76426; AAC42061.1; -.
DR   EMBL; M76427; AAC42062.1; -.
DR   MEROPS; S09.973; -.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
KW   Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
FT   DOMAIN        1     89       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     90    110       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN      111    859       EXTRACELLULAR (POTENTIAL).
FT   CARBOHYD    167    167       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    168    168       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    313    313       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    398    398       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    465    465       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    529    529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    560    560       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    807    807       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC      1     75       MASLYQRFTGKINTSRSFPAPPEASHLLGGQGPEEDAGSKP
FT                                LGPQAQAVAPRERGGAGGRPRFQYQARSDCDEED -> MTT
FT                                AKEPSASGKSVQQQDQ (in isoform DPPX-S).
FT                                /FTId=VSP_005366.
SQ   SEQUENCE   859 AA;  97301 MW;  CE26856D26ED126B CRC64;

  Query Match             11.0%;  Score 212;  DB 1;  Length 859;
  Best Local Similarity   19.7%;  Pred. No. 2.7e-10;
  Matches   98;  Conservative   67;  Mismatches  158;  Indels  174;  Gaps   16;

Qy          8 GVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV---TSPMLE 63
              |:: :: :||    :  :|| |             |: |   ::| | :: :   |  :
Db        281 GLSDWLYEEEILKSHIAHWWSPDG---------TRLAYATINDSRVPLMELPTYTGSVYP 331

Qy         64 TRRADSFRYPKTGTANPKVTF-------------------KMSEIMI------------ 91
```

```
              |     : ||| |: || ::                        :| |  |
Db        332 T--VKPYHYPKAGSENPSISLHVIGLNGPTHDLEMMPPDDPRMREYYITMVKWATSTKVA 389

Qy         92 ----------------DA----------------------------EGR----------- 96
                            ||                            :||
Db        390 VTWLNRAQNVSILTLCDATTGVCTKKHEDESEAWLHRQNEEPVFSKDGRKFFFVRAIPQG 449

Qy         97 -----------------------------------IIVDEVRRLVYFEGTKDSPLEHHLY 121
                                                 :  || |  :||  |:| |   |||
Db        450 GRGKFYHITVSSSQPNSSNDNIQSITSGDWDVTEILTYDEKRNKLYFLSTEDLPRRRHLY 509

Qy        122 VVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSP------ED 175
               : |:      |      || : ::| :  : :|:  :     | |   |
Db        510 SANTVDDFNRQCL-------SCDLVENCTYVSASFSHNMD---FFLLKCEGPGVPTVTVH 559

Qy        176 DPTCKTKEF--------WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPG 227
               : | : |           | |   |  :|   |:        ::|   : ||
Db        560 NTTDKRRMFDLEANEQVQKAIYDRQMPKIEYRKIEV------EDYSLPMQILKPATFTDT 613

Qy        228 KKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFK 287
               || :| : | |  | |: ||:    :    ::| | ||| | ||| :| |     :
Db        614 AHYPLLLVVDGTPGSQSVSERFEVT--WETVLVSSHGAVVVKCDGRGSGFQGTKLLHEVR 671

Qy        288 YKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSD----IFRVA 343
              ::| :|  ||:| :: :  :  :||  || : |  ||||||   |  : :     |
Db        672 RRLGPLEEKDQMEAVRTML-KEQYIDKTRVAVFGKDYGGYLSTYILPAKGENQGQTFTCG 730

Qy        344 IAGAPVTLWIFYDTGYT 360
               | :|:| :  | : ::
Db        731 SALSPITDFKLYASAFS 747
```

```
RESULT 13
DPP6_MOUSE
ID   DPP6_MOUSE     STANDARD;      PRT;   804 AA.
AC   Q9Z218; Q9QWW2; Q9Z219;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
DE   (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
DE   (Dipeptidyl aminopeptidase-related protein) (DPPX).
GN   DPP6 OR DPP-6.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Embryo;
RX   MEDLINE=99030650; PubMed=9811881;
RA   Hough R.B., Lengeling A., Bedian V., Lo C., Bucan M.;
RT   "Rump white inversion in the mouse disrupts dipeptidyl aminopeptidase-
RT   like protein 6 and causes dysregulation of kit expression.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:13800-13805(1998).
CC   -!- FUNCTION: May be involved in the physiological processes of brain
CC       function. Has no dipeptidyl aminopeptidase activity. The lack of
CC       activity may be due to the substitution of an aspartate residue
CC       for the serine residue in the proposed catalytic triad (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF092507; AAC97366.1; -.
DR   EMBL; AF092506; AAC97365.1; -.
```

```
DR   EMBL; AF092505; AAC98381.1; -.
DR   MEROPS; S09.973; -.
DR   MGD; MGI:94921; Dpp6.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
KW   Transmembrane; Glycoprotein; Signal-anchor.
FT   DOMAIN        1     34       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     35     55       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       56    804       EXTRACELLULAR (POTENTIAL).
FT   CARBOHYD    112    112       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    113    113       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    258    258       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    343    343       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    410    410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    474    474       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    505    505       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    752    752       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    440    455       MISSING (IN REF. 1; AAC98381).
FT   CONFLICT    638    638       Q -> P (IN REF. 1; AAC97365).
SQ   SEQUENCE   804 AA;  91260 MW;  09CFCE7AAD8A7168 CRC64;

  Query Match             10.9%;  Score 211;  DB 1;  Length 804;
  Best Local Similarity   19.7%;  Pred. No. 3e-10;
  Matches   98;  Conservative   65;  Mismatches  159;  Indels  176;  Gaps   16;

Qy          8 GVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV---TSPMLE 63
              |:: :: :||    :  :|| |          |: |   ::| | :: :   |  :
Db        226 GLSDWLYEEEILKSHIAHWWSPDG---------TRLAYATINDSRVPLMELPTYTGSVYP 276

Qy         64 TRRADSFRYPKTGTANPKVTF--------------------KMSEIMI------------ 91
              |     : ||| |: || ::                     :| |  |
Db        277 T--VKPYHYPKAGSENPSISLHVIGLNGPTHDLEMMPPDDPRMREYYITMVKWATSTKVA 334

Qy         92 ----------------DA----------------------------EGR----------- 96
                            ||                            :||
Db        335 VTWLNRAQNVSILTLCDATTGVCTKKHEDESEAWLHRQNEEPVFSKDGRKFFFVRAIPQG 394

Qy         97 ---------------------------------IIVDEVRRLVYFEGTKDSPLEHHLY 121
                                               :  || |  :||  |:| |   |||
Db        395 GRGKFYHITVSSSQPNSSNDNIQSITSGDWDVTKILSYDEKRNKIYFLSTEDLPRRRHLY 454

Qy        122 VVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTC-- 179
              : |:       |       || : ::| :  : :|:  :     | |   |   ||
Db        455 SANTVDDFNRQCL-------SCDLVENCTYVSASFSHNMD---FFLLKCEGP-GVPTVTV 503

Qy        180 ------------KTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQP 226
                         :|    | |   |  :|   |:        ::|   : ||
Db        504 HNTTDKRRMFDLEANEEVQKAINDRQMPKIEYRKIEV------EDYSLPMQILKPATFTD 557

Qy        227 GKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAF 286
              || :| : | |  | |  ||:    :    ::| | |||  | |||  :| |
Db        558 TAHYPLLLVVDGTPGSQSVTERFEVT--WETVLVSSHGAVVVKCDGRGSGFQGTKLLQEV 615

Qy        287 KYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSD----IFRV 342
              : ::| :|  ||:| :: :  :  :||  || : |  ||||||   |  : :     |
Db        616 RRRLGFLEEKDQMEAVRTML-KEQYIDKTRVAVFGKDYGGYLSTYILPAKGENQGQTFTC 674

Qy        343 AIAGAPVTLWIFYDTGYT 360
              | :|:| :  | : ::
Db        675 GSALSPITDFKLYASAFS 692


RESULT 14
DPP6_HUMAN
ID   DPP6_HUMAN     STANDARD;      PRT;   865 AA.
AC   P42658;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
DE   (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
DE   (Dipeptidyl aminopeptidase-related protein) (DPPX).
GN   DPP6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S).
RC   TISSUE=Hippocampus;
RX   MEDLINE=93372805; PubMed=8103397;
RA   Yokotani N., Doi K., Wenthold R.J., Wada K.;
RT   "Non-conservation of a catalytic residue in a dipeptidyl
RT   aminopeptidase IV-related protein encoded by a gene on human
RT   chromosome 7.";
RL   Hum. Mol. Genet. 2:1037-1039(1993).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
CC       FUNCTION. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
CC       ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
CC       FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=DPPX-L;
CC         IsoId=P42658-1; Sequence=Displayed;
CC       Name=DPPX-S;
CC         IsoId=P42658-2; Sequence=VSP_005365;
CC   -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M96859; AAA35760.1; -.
DR   EMBL; M96860; AAA35761.1; -.
DR   PIR; I54331; I54331.
DR   MEROPS; S09.973; -.
DR   Genew; HGNC:3010; DPP6.
DR   MIM; 126141; -.
DR   GO; GO:0008239; F:dipeptidyl-peptidase activity; TAS.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
KW   Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
FT   DOMAIN        1     95       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     96    116       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       117    865       EXTRACELLULAR (POTENTIAL).
FT   CARBOHYD     173    173       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     319    319       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     404    404       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     471    471       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     535    535       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     566    566       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     813    813       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC       1     81       MASLYQRFTGKINTSRSFPAPPEASHLLGGQGPEEDGGAGA
FT                                KPLGPRAQAAAPRERGGGGGGAGGRPRFQYQGRSDGDEED
FT                                -> MTTAKEPSASGKSVQQQEQE (in isoform
FT                                DPPX-S).
FT                                /FTId=VSP_005365.
SQ   SEQUENCE   865 AA;  97588 MW;  14B1AE0E0024464B CRC64;

  Query Match             10.5%;  Score 203;  DB 1;  Length 865;
  Best Local Similarity   19.6%;  Pred. No. 1.6e-09;
  Matches   98;  Conservative   65;  Mismatches  157;  Indels  180;  Gaps   17;
```

```
Qy          8 GVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV---TSPMLE 63
              |:: :: :||    :  :|| |            |: |   ::| | |: :   |  :
Db        287 GLSDWLYEEEILKTHIAHWWSPDG---------TRLAYAAINDSRVPIMELPTYTGSIYP 337

Qy         64 TRRADSFRYPKTGTANPKVTF--------------------KMSEIMI------------ 91
              |    : ||| |: || ::                      :| | |
Db        338 T--VKPYHYPKAGSENPSISLHVIGLNGPTHDLEMMPPDDPRMREYYITMVKWATSTKVA 395

Qy         92 ----------------DA---------------------------EGR----------- 96
                            ||                           :||
Db        396 VTWLNRAQNVSILTLCDATTGVCTKKHEDESEAWLHRQNEEPVFSKDGRKFFFIRAIPQG 455

Qy         97 ----------------------------------IIVDEVRRLVYFEGTKDSPLEHHLY 121
                                               :  ||    :||  |:| |    ||
Db        456 GRGKFYHITVSSSQPNSSNDNIQSITSGDWDVTKILAYDEKGNKIYFLSTEDLPRRRQLY 515

Qy        122 VVSYVNPGEVTRLTDRGYSH---SCCISQHCDFFISKYSNQKNPHCVSLYKLSSP----- 173
                :          |:  ::    || : ::| :| : :|:  :     | |   |
Db        516 SAN----------TEGNFNRQCLSCDLVENCTYFSASFSHSMD---FFLLKCEGPGVPMV 562

Qy        174 -EDDPTCKTKEF--------WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDL 224
               : | | | |           | |   |  :|   ||       : |   : ||
Db        563 TVHNTTDKKKMFDLETNEHVKKAINDRQMPKVEYRDIEI------DDYNLPMQILKPATF 616

Qy        225 QPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEG 284
                || :| : | |  | |  :|:    :    ::| | ||| | ||| :| |
Db        617 TDTTHYPLLLVVDGTPGSQSVAEKFE--VSWETVMVSSHGAVVVKCDGRGSGFQGTKLLH 674

Qy        285 AFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSD----IF 340
               : ::| :|  ||:| :: :  :  :||  || : |  ||||||   |  : :     |
Db        675 EVRRRLGLLEEKDQMEAVRTML-KEQYIDRTRVAVFGKDYGGYLSTYILPAKGENQGQTF 733

Qy        341 RVAIAGAPVTLWIFYDTGYT 360
                | :|:| :. | : ::
Db        734 TCGSALSPITDFKLYASAFS 753
```

```
RESULT 15
YDZF_SCHPO
ID   YDZF_SCHPO     STANDARD;      PRT;   853 AA.
AC   Q9P7E9; O13720;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Putative dipeptidyl aminopeptidase C14C4.15c (EC 3.4.14.-).
GN   SPAC14C4.15C OR SPAPJ760.01C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like
CC       vacuoles.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AL162631; CAB83084.1; -.
DR   EMBL; Z98596; CAB11208.1; -.
DR   GeneDB_SPombe; SPAC14C4.15c; -.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
KW   Hypothetical protein; Hydrolase; Aminopeptidase; Dipeptidase;
KW   Serine protease; Transmembrane; Glycoprotein; Signal-anchor.
FT   DOMAIN        1     66       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     67     89       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       90    853       LUMENAL (POTENTIAL).
FT   ACT_SITE    719    719       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    795    795       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    828    828       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     96     96       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    102    102       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    472    472       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    613    613       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   853 AA;  98341 MW;  38450BA50F8304B6 CRC64;

  Query Match             10.2%;  Score 197.5;  DB 1;  Length 853;
  Best Local Similarity   29.8%;  Pred. No. 4.7e-09;
  Matches   53;  Conservative   25;  Mismatches   59;  Indels   41;  Gaps    6;

Qy        200 PPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNT 259
              |   |         | |    :| :  | |:|||| :|| ||  ||  :::      :|
Db        602 PSVFFKVIKVKNITAYIQEIRPPNFNPRKRYPTVFHLYGAPQSALVTGKYE----MDINE 657

Qy        260 LASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEG--------------LQY 304
              | :  |  :||                  |:  ::| | | |              ::
Db        658 LMASVYNFLVI-----------------KVDIVDISD-VSGQHLFSDSHELIIKSWIEL 698

Qy        305 LASRYD--FIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              | |  |  :||  |||| |||:||||:|  ::: :|  :     |||| | :||  |:
Db        699 LRSYVDTPYIDRHRVGIWGWSFGGYLTL-KILENADFIKTGAVVAPVTDWRYYDAYYS 755


Search completed: October 15, 2003, 17:11:34
Job time : 7.71145 secs
```

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:02:16 ; Search time 23.0243 Seconds
(without alignments)
4034.822 Million cell updates/sec

Title: US-10-070-464-7
Perfect score: 1933
Sequence: 1 EEDARSAGVATFVLQEEFDR.....RVAIAGAPVTLWIFYDTGYT 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1933	100.0	360	4	Q9HBM2	Q9hbm2 homo sapien
2	1836.5	95.0	882	4	Q9HBM5	Q9hbm5 homo sapien
3	1836.5	95.0	882	4	Q8IWG7	Q8iwg7 homo sapien
4	1756.5	90.9	892	11	Q9D4G6	Q9d4g6 mus musculu
5	1534	79.4	831	4	Q8NEM5	Q8nem5 homo sapien
6	1528	79.0	632	4	Q96JX1	Q96jx1 homo sapien
7	1220.5	63.1	628	4	Q8N3F5	Q8n3f5 homo sapien
8	1220.5	63.1	863	4	Q8WXD8	Q8wxd8 homo sapien
9	1204.5	62.3	862	11	Q8BWT9	Q8bwt9 mus musculu
10	1204.5	62.3	862	11	Q8BVG4	Q8bvg4 mus musculu
11	1086.5	56.2	310	4	Q9HBM4	Q9hbm4 homo sapien
12	984	50.9	439	4	Q9BVR3	Q9bvr3 homo sapien
13	944	48.8	508	4	O75273	O75273 homo sapien
14	860.5	44.5	465	4	Q9HBM3	Q9hbm3 homo sapien
15	799	41.3	312	4	Q96NT8	Q96nt8 homo sapien
16	722	37.4	1042	5	Q9VC20	Q9vc20 drosophila
17	722	37.4	1102	5	Q9VC19	Q9vc19 drosophila
18	718.5	37.2	1053	5	Q8IH07	Q8ih07 drosophila
19	527.5	27.3	738	16	Q9A6E0	Q9a6e0 caulobacter
20	522	27.0	469	4	Q9NXF4	Q9nxf4 homo sapien
21	494	25.6	763	16	Q8EAB7	Q8eab7 shewanella
22	491.5	25.4	927	5	Q965K3	Q965k3 caenorhabdi
23	491.5	25.4	931	5	O44987	O44987 caenorhabdi
24	471	24.4	751	16	Q8P3V8	Q8p3v8 xanthomonas
25	463	24.0	757	16	Q8PFD7	Q8pfd7 xanthomonas
26	430.5	22.3	741	2	P95782	P95782 xanthomonas
27	422.5	21.9	746	10	Q9FNF6	Q9fnf6 arabidopsis
28	390.5	20.2	552	10	Q8GUJ7	Q8guj7 arabidopsis
29	369	19.1	711	2	Q47900	Q47900 flavobacter
30	366.5	19.0	730	2	Q93JY4	Q93jy4 prevotella
31	347.5	18.0	771	3	O42812	O42812 aspergillus
32	346	17.9	723	2	O31048	O31048 porphyromon
33	344	17.8	723	2	O66223	O66223 porphyromon
34	343.5	17.8	432	4	O75868	O75868 homo sapien
35	322	16.7	237	2	O87543	O87543 capnocytoph
36	320	16.6	765	3	O14425	O14425 aspergillus
37	319	16.5	827	16	Q8PPU4	Q8ppu4 xanthomonas
38	317.5	16.4	748	13	P70092	P70092 xenopus lae
39	313.5	16.2	901	3	Q96VT7	Q96vt7 aspergillus
40	304.5	15.8	707	16	Q9F348	Q9f348 streptomyce
41	291	15.1	803	16	Q8P500	Q8p500 xanthomonas
42	288	14.9	167	4	Q8N2J7	Q8n2j7 homo sapien
43	282	14.6	935	5	Q9VMB4	Q9vmb4 drosophila
44	281	14.5	745	5	Q9VMM2	Q9vmm2 drosophila
45	271	14.0	802	5	Q9VUH1	Q9vuh1 drosophila

ALIGNMENTS

RESULT 1
Q9HBM2
ID Q9HBM2 PRELIMINARY; PRT; 360 AA.
AC Q9HBM2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Dipeptidyl peptidase 8 (Fragment).
GN DPP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
RA Gorrell M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
RT dipeptidyl peptidase (DPP) IV homolog, DPP8.";
RL Eur. J. Biochem. 267:6140-6150(2000).
DR EMBL; AF221637; AAG29769.1; -.
DR MEROPS; S09.018; -.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00326; Peptidase_S9; 1.
FT NON_TER 1 1
FT NON_TER 360 360
SQ SEQUENCE 360 AA; 41070 MW; CF81C0BB61423E22 CRC64;

Query Match 100.0%; Score 1933; DB 4; Length 360;
Best Local Similarity 100.0%; Pred. No. 3.7e-168;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
Db 1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60

Qy 61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIVDEVRRLVYFEGTKDSPLEHHL 120

```
Db         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIVDEVRRLVYFEGTKDSPLEHHL 120

Qy        121 YVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCK 180

Db        121 YVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCK 180

Qy        181 TKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGP 240

Db        181 TKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGP 240

Qy        241 QVQLVNNRFKGVKYFRLNTLASLGYVVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVE 300

Db        241 QVQLVNNRFKGVKYFRLNTLASLGYVVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVE 300

Qy        301 GLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360

Db        301 GLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360


RESULT 2
Q9HBM5
ID   Q9HBM5      PRELIMINARY;      PRT;   882 AA.
AC   Q9HBM5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Dipeptidyl peptidase 8.
GN   DPP8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20467194; PubMed=11012666;
RA   Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
RA   Gorrell M.D.;
RT   "Cloning, expression and chromosomal localization of a novel human
RT   dipeptidyl peptidase (DPP) IV homolog, DPP8.";
RL   Eur. J. Biochem. 267:6140-6150(2000).
DR   EMBL; AF221634; AAG29766.1; -.
DR   MEROPS; S09.018; -.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
SQ   SEQUENCE   882 AA;  101421 MW;  AD801C302DB4652B CRC64;

  Query Match             95.0%;  Score 1836.5;  DB 4;  Length 882;
  Best Local Similarity   67.5%;  Pred. No. 8.4e-159;
  Matches  360;  Conservative    0;  Mismatches    0;  Indels  173;  Gaps    1;

Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60

Db        244 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 303

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98

Db        304 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 363

Qy         99 ------------------------------------------------------------ 98

Db        364 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 423

Qy         99 ------------------------------------------------------------ 98

Db        424 EETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 483

Qy         99 ----------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127

Db        484 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 543

Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187

Db        544 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 603

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247

Db        604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 663

Qy        248 RFKGVKYFRLNTLASLGYVVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307

Db        664 RFKGVKYFRLNTLASLGYVVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 723

Qy        308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360

Db        724 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 776


RESULT 3
Q8IWG7
ID   Q8IWG7      PRELIMINARY;      PRT;   882 AA.
AC   Q8IWG7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to dipeptidylpeptidase 8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Strausberg R.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC040203; AAH40203.1; -.
SQ   SEQUENCE   882 AA;  101391 MW;  88C76AF5BCE707F9 CRC64;

  Query Match             95.0%;  Score 1836.5;  DB 4;  Length 882;
  Best Local Similarity   67.5%;  Pred. No. 8.4e-159;
  Matches  360;  Conservative    0;  Mismatches    0;  Indels  173;  Gaps    1;

Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60

Db        244 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 303

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98

Db        304 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 363

Qy         99 ------------------------------------------------------------ 98

Db        364 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 423

Qy         99 ------------------------------------------------------------ 98

Db        424 EETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 483

Qy         99 ----------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127

Db        484 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 543

Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187

Db        544 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 603

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247

Db        604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 663

Qy        248 RFKGVKYFRLNTLASLGYVVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307

Db        664 RFKGVKYFRLNTLASLGYVVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 723
```

```
Qy        308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 776
```

RESULT 4
Q9D4G6
ID Q9D4G6 PRELIMINARY; PRT; 892 AA.
AC Q9D4G6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Adult male testis cDNA, RIKEN full-length enriched library,
DE clone:4932434F09 product:DIPEPTIDYL PEPTIDASE 8 homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
DR EMBL; AK016546; BAB30295.2; -.
SQ SEQUENCE 892 AA; 102284 MW; FDE0DBEDCA4CA346 CRC64;

  Query Match 90.9%; Score 1756.5; DB 11; Length 892;
  Best Local Similarity 64.4%; Pred. No. 1.7e-151;
  Matches 343; Conservative 7; Mismatches 10; Indels 173; Gaps 1;

```
Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              ||| |||||||||||||||||||||||:|| ||||||||||||||||||||||||||||
Db        254 EEDPRSAGVATFVLQEEFDRYSGYWWCPQAERTPSGGKILRILYEENDESEVEIIHVTSP 313

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              ||||||||||||||||||||||||||||::|| | ||
Db        314 MLETRRADSFRYPKTGTANPKVTFKMSEIVVDAAGGIIDVIDKELVQPFEILFEGVEYIA 373

Qy         99 ---------------------------------------------------------- 98

Db        374 RAGWTPEGKHAWSILLDRSQTHLQIVLISPELFIPVEDDAMDRQRLIESVPDSVTPLIIY 433

Qy         99 ---------------------------------------------------------- 98

Db        434 EETTDIWINIHDIFHVFPQTHEDEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 493

Qy         99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                            ||| |:|||||||||||||||| || |
Db        494 APSDFKCPIKEEITITSGEWEVLGRHGSNIWVDEARKLVYFEGTKDSPLEHHLYVTSYAN 553

Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
              |||| |||||||||||:|:|||||||||||||||||||||||||||||  ||||||||
Db        554 PGEVVRLTDRGYSHSCCLSRHCDFFISKYSNQKNPHCVSLYKLSSPEDDPVHKTKEFWAT 613

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        614 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 673

Qy        248 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        674 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 733

Qy        308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              :||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        734 QYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 786
```

RESULT 5
Q8NEM5
ID Q8NEM5 PRELIMINARY; PRT; 831 AA.
AC Q8NEM5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Similar to dipeptidylpeptidase 8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC030688; AAH30688.1; -.
SQ SEQUENCE 831 AA; 95527 MW; 0B2A13A2FE70CBE2 CRC64;

  Query Match 79.4%; Score 1534; DB 4; Length 831;
  Best Local Similarity 58.0%; Pred. No. 3.4e-131;

```
  Matches  309;  Conservative    0;  Mismatches    0;  Indels  224;  Gaps    2;

Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 303

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              ||||||||||||||||||||||||||||||||||||
Db        304 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 363

Qy         99 ------------------------------------------------------------ 98

Db        364 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 423

Qy         99 ------------------------------------------------------------ 98

Db        424 EETTDIWINIHDIFHVFPQSHEEEIEFIPASECKTGFRHLYKITSILKESKYKRSSGGLP 483

Qy         99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                           ||||||||||||||||||||||||||||||
Db        484 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 543

Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 603

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 663

Qy        248 RFKGVKYFRLNTLASLGYVVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307
              |||||||||||||||||||||||||||||||||||||||||||
Db        664 RFKGVKYFRLNTLASLGYVVVVVIDNRGSCHRGLKFEGAFKYKM---------------- 706

Qy        308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
                                                 ||||||||||||||||||
Db        707 ---------------------------------VAIAGAPVTLWIFYDTGYT 725
```

RESULT 6
Q96JX1
ID Q96JX1 PRELIMINARY; PRT; 632 AA.
AC Q96JX1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ14920.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK027826; BAB55395.1; -.
DR MEROPS; S09.018; -.
KW Hypothetical protein.
SQ SEQUENCE 632 AA; 72639 MW; 9BDF598B06985AA4 CRC64;

```
  Query Match           79.0%;  Score 1528;  DB 4;  Length 632;
  Best Local Similarity  57.8%;  Pred. No. 8.2e-131;
  Matches  308;  Conservative    0;  Mismatches    1;  Indels  224;  Gaps    2;

Qy          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         45 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 104

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              ||||||||||||||||||||||||||||||||||||
Db        105 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 164

Qy         99 ------------------------------------------------------------ 98

Db        165 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 224

Qy         99 ------------------------------------------------------------ 98

Db        225 EETTDIWINIHDIFHVFPQSHEEEIEFIPASECKTGFRHLYKITSILKESKYKRSSGGLP 284

Qy         99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127
                                           ||||||||||||||||||||||||||||||
Db        285 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 344

Qy        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        345 PGEVTRLTDRSYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 404

Qy        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        405 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 464

Qy        248 RFKGVKYFRLNTLASLGYVVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307
              |||||||||||||||||||||||||||||||||||||||||||
Db        465 RFKGVKYFRLNTLASLGYVVVVVIDNRGSCHRGLKFEGAFKYKM---------------- 507

Qy        308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
                                                 ||||||||||||||||||
Db        508 ---------------------------------VAIAGAPVTLWIFYDTGYT 526
```

RESULT 7
Q8N3F5
ID Q8N3F5 PRELIMINARY; PRT; 628 AA.
AC Q8N3F5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP762M2413.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RA Ansorge W., Wirkner U., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL834376; CAD39039.1; -.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00326; Peptidase_S9; 1.
KW Hypothetical protein.
FT NON_TER 1 1
SQ SEQUENCE 628 AA; 71368 MW; DC0B6A9440507C3B CRC64;

```
  Query Match           63.1%;  Score 1220.5;  DB 4;  Length 628;
  Best Local Similarity  44.2%;  Pred. No. 1e-102;
  Matches  236;  Conservative   54;  Mismatches   69;  Indels  175;  Gaps    2;

Qy          2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
              :| :||||||||:||||||::|||||| |    | | | ||||||| ||||||:||| ||
Db         26 DDPKSAGVATFVIQEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSP 85

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
              || |: ||:|||:||: |||:  |::|  |::|:|:
Db         86 ALEERKTDSYRYPRTGSKNPKIALKLAEFQTDSQGKIVSTQEKELVQPFSSLFPKVEYIA 145
```

```
Qy         99 ------------------------------------------------------------ 98

Db        146 RAGWTRDGKYAWAMFLDRPQQWLQLVLLPPALFIPSTENEEQRLASARAVPRNVQPYVVY 205

Qy         99 ------------------------------------------------------------ 98

Db        206 EEVTNVWINVHDIFYPFPNQRERTSSAFSAPMNARPASAICTKSPPFLKSQGYDWSEPFS 265

Qy         99 --------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYV 126
                                              |:|  :||||:||||:||||||||||
Db        266 PGEDEFKCPIKEEIALTSGEWEVLARHGSKIWVNEETKLVYFQGTKDTPLEHHLYVVSYE 325

Qy        127 NPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWA 186
                 ||: |||  |:||||  :||: | |:| ||:    | || :|||| |:|||  |    |||
Db        326 AAGEIVRLTTPGFSHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWA 385

Qy        187 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVN 246
              :::::|    ||| ||||| | : :    ||||:|||| ||||||:||||||:|||||||||
Db        386 SMMEAASCPPDYVPPEIFHFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVN 445

Qy        247 NRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLA 306
              | |||:|| |||||||||| ||||| |||| |||:||||  | :|||:|||||||:|:|
Db        446 NSFKGIKYLRLNTLASLGYAVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVA 505

Qy        307 SRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
               :| |||| || ||||||||:|||| |: :   :|:|||||||||:|: ||||||
Db        506 EKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVWMAYDTGYT 559
```

RESULT 8
Q8WXD8
```
ID   Q8WXD8      PRELIMINARY;      PRT;   863 AA.
AC   Q8WXD8;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Dipeptidyl peptidase 9.
GN   DPP9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Olsen C., Wagtmann N.;
RT   "Identification and characterization of a novel member of the
RT   dipeptidyl peptidase IV-related family.";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RA   Strausberg R.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF452102; AAL47179.1; -.
DR   EMBL; BC037948; AAH37948.1; -.
DR   MEROPS; S09.019; -.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
SQ   SEQUENCE   863 AA;  98263 MW;  40FE0B78E26CDED5 CRC64;

  Query Match             63.1%;  Score 1220.5;  DB 4;  Length 863;
  Best Local Similarity   44.2%;  Pred. No. 1.6e-102;
  Matches  236;  Conservative   54;  Mismatches   69;  Indels  175;  Gaps    2;

Qy          2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
                :| :||||||||:||||||::||||||  |     | | | ||||||| ||||||:||| ||
Db        234 DDPKSAGVATFVIQEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSP 293

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
```

```
               || |: ||:|||:||: |||:  |::|  |::|:|:
Db        294 ALEERKTDSYRYPRTGSKNPKIALKLAEFQTDSQGKIVSTQEKELVQPFSSLFPKVEYIA 353

Qy         99 ------------------------------------------------------------ 98

Db        354 RAGWTRDGKYAWAMFLDRPQQWLQLVLLPPALFIPSTENEEQRLASARAVPRNVQPYVVY 413

Qy         99 ------------------------------------------------------------ 98

Db        414 EEVTNVWINVHDIFYPFPQSEGEDELCFLRANECKTGFCHLYKVTAVLKSQGYDWSEPFS 473

Qy         99 --------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYV 126
                                              |:|  :||||:||||:||||||||||
Db        474 PGEDEFKCPIKEEIALTSGEWEVLARHGSKIWVNEETKLVYFQGTKDTPLEHHLYVVSYE 533

Qy        127 NPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWA 186
                 ||: |||  |:||||  :||: | |:| ||:    | || :|||| |:|||  |    |||
Db        534 AAGEIVRLTTPGFSHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWA 593

Qy        187 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVN 246
              :::::|    ||| ||||| | : :    ||||:|||| ||||||:||||||:|||||||||
Db        594 SMMEAASCPPDYVPPEIFHFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVN 653

Qy        247 NRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLA 306
              | |||:|| |||||||||| ||||| |||| |||:||||  | :|||:|||||||:|:|
Db        654 NSFKGIKYLRLNTLASLGYAVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVA 713

Qy        307 SRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
               :| |||| || ||||||||:|||| |: :   :|:|||||||||:|: ||||||
Db        714 EKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVWMAYDTGYT 767
```

RESULT 9
Q8BWT9
```
ID   Q8BWT9      PRELIMINARY;      PRT;   862 AA.
AC   Q8BWT9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Dipeptidyl peptidase 9 homolog.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Liver;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK050021; BAC34034.1; -.
SQ   SEQUENCE   862 AA;  98001 MW;  B1D566E824A834E8 CRC64;

  Query Match             62.3%;  Score 1204.5;  DB 11;  Length 862;
  Best Local Similarity   43.8%;  Pred. No. 4.6e-101;
  Matches  234;  Conservative   53;  Mismatches   72;  Indels  175;  Gaps    2;

Qy          2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
                :: :||||||||:||||||::| ||||  |     | | | ||||||| ||||||:||| ||
Db        233 DNPKSAGVATFVIQEEFDRFTGCWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSP 292

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
               || |: ||:|||:||: |||:  |::|: | :|:|:
Db        293 ALEERKTDSYRYPRTGSKNPKIALKLAELQTDHQGKIVSSCEKELVQPFSSLFPKVEYIA 352

Qy         99 ------------------------------------------------------------ 98

Db        353 RAGWTRDGKYAWAMFLDRPQQRLQLVLLPPALFIPAVESEAQRQAAARAVPKNVQPFVIY 412

Qy         99 ------------------------------------------------------------ 98
```

```
Db        413 EEVTNVWINVHDIFHPFPQAEGQQDFCFLRANECKTGFCHLYRVTVELKTKDYDWTEPLS 472

Qy         99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYV 126
                                            |:|  :||||:||||:||||||||||||
Db        473 PTEDEFKCPIKEEVALTSGEWEVLSRHGSKIWVNEQTKLVYFQGTKDTPLEHHLYVVSYE 532

Qy        127 NPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWA 186
              : ||: |||  |:|||| :||  | |:|  ||:   | || :|||| |:|||  |    |||
Db        533 SAGEIVRLTTLGFSHSCSMSQSFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWA 592

Qy        187 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVN 246
              :::::|   ||| ||||| | :     ||||:|||| ||||:|:||||:|:||||||||||
Db        593 SMMEAANCPPDYVPPEIFHFHTRADVQLYGMIYKPHTLQPGRKHPTVLFVYGGPQVQLVN 652

Qy        247 NRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLA 306
              | |||:|| |||||||||| ||||| |||| ||| |||| | :|||:||:|||||||:|
Db        653 NSFKGIKYLRLNTLASLGYAVVVIDGRGSCQRGLHFEGALKNQMGQVEIEDQVEGLQYVA 712

Qy        307 SRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
                :| |||| || |||||||||:|||| |: :  :|:||||||||||:|: ||||||
Db        713 EKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVWMAYDTGYT 766


RESULT 10
Q8BVG4
ID   Q8BVG4      PRELIMINARY;      PRT;   862 AA.
AC   Q8BVG4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Dipeptidyl peptidase 9 homolog.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK078301; BAC37211.1; -.
SQ   SEQUENCE   862 AA;  98049 MW;  B4FD3476B6F50030 CRC64;

  Query Match             62.3%;  Score 1204.5;  DB 11;  Length 862;
  Best Local Similarity   43.8%;  Pred. No. 4.6e-101;
  Matches  234;  Conservative   53;  Mismatches   72;  Indels . 175;  Gaps    2;

Qy          2 EDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGG-KILRILYEENDESEVEIIHVTSP 60
              :: :|||||||:||||||::| |||| |     | | ||||||| |||||||||| ||
Db        233 DNPKSAGVATFVIQEEFDRFTGCWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSP 292

Qy         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98
                 || |: ||:|||:||: |||:  |::|:  | :|:|:
Db        293 ALEERKTDSYRYPRTGSKNPKIALKLAELQTDHQGKIVSSCEKELVQPFSSLFPKVEYIA 352

Qy         99 ---------------------------------------------------------- 98

Db        353 RAGWTRDGKYAWAMFLYRPQQRLQLVLLPPALFIPAVESEAQRQAAARAVPKNVQPFVIY 412

Qy         99 ---------------------------------------------------------- 98

Db        413 EEVTNVWINVHDIFHPFPQAEGQQDFCFLRANECKTGFCHLYRVTVELKTKDYDWTEPLS 472

Qy         99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYV 126
                                            |:|  :||||:||||:||||||||||||
Db        473 PTEDEFKCPIKEEVALTSGEWEVLSRHGSKIWVNEQTKLVYFQGTKDTPLEHHLYVVSYE 532

Qy        127 NPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWA 186
              : ||: |||  |:|||| :||  | |:|  ||:   | || :|||| |:|||  |    |||
Db        533 SAGEIVRLTTLGFSHSCSMSQSFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWA 592

Qy        187 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVN 246
              :::::|   ||| ||||| | :     ||||:|||| ||||:|:||||:|:||||||||||
Db        593 SMMEAANCPPDYVPPEIFHFHTRADVQLYGMIYKPHTLQPGRKHPTVLFVYGGPQVQLVN 652

Qy        247 NRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLA 306
              | |||:|| |||||||||| ||||| |||| ||| |||| | :|||:||:|||||||:|
Db        653 NSFKGIKYLRLNTLASLGYAVVVIDGRGSCQRGLHFEGALKNQMGQVEIEDQVEGLQYVA 712

Qy        307 SRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
                :| |||| || |||||||||:|||| |: :  :|:||||||||||:|: ||||||
Db        713 EKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVWMAYDTGYT 766


RESULT 11
Q9HBM4
ID   Q9HBM4      PRELIMINARY;      PRT;   310 AA.
AC   Q9HBM4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Dipeptidyl peptidase 8 (Fragment).
GN   DPP8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=20467194; PubMed=11012666;
RA   Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
RA   Gorrell M.D.;
RT   "Cloning, expression and chromosomal localization of a novel human
RT   dipeptidyl peptidase (DPP) IV homolog, DPP8.";
RL   Eur. J. Biochem. 267:6140-6150(2000).
DR   EMBL; AF221635; AAG29767.1; -.
DR   MEROPS; S09.018; -.
DR   InterPro; IPR001375; Peptidase_S9.
DR   Pfam; PF00326; Peptidase_S9; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   310 AA;  35396 MW;  8E87C34026D9C7AC CRC64;

  Query Match             56.2%;  Score 1086.5;  DB 4;  Length 310;
  Best Local Similarity   80.6%;  Pred. No. 6.7e-91;
  Matches  204;  Conservative    0;  Mismatches    0;  Indels   49;  Gaps    1;

Qy        108 FEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSL 167
              o ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 FEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSL 60

Qy        168 YKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPG 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPG 120

Qy        228 KKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFK 287
              ||||||||||||||
Db        121 KKYPTVLFIYGGPQ----------------------------------------------- 134

Qy        288 YKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGA 347
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 ---GQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGA 191

Qy        348 PVTLWIFYDTGYT 360
              |||||||||||||
Db        192 PVTLWIFYDTGYT 204


RESULT 12
Q9BVR3
ID   Q9BVR3      PRELIMINARY;      PRT;   439 AA.
```

```
AC   Q9BVR3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Strausberg R.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC000970; AAH00970.1; -.
DR   MEROPS; S09.019; -.
DR   Genew; HGNC:18648; DPP9.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00326; Peptidase_S9; 1.
KW   Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   439 AA;  49926 MW;  A18BBA9E12092BAF CRC64;

  Query Match             50.9%;  Score 984;  DB 4;  Length 439;
  Best Local Similarity   64.8%;  Pred. No. 2.5e-81;
  Matches  177;  Conservative   40;  Mismatches   56;  Indels    0;  Gaps    0;

Qy         88 EIMIDAEGRIIVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQ 147
              |::     :| |:|  :||||:||||:||||||||||    ||: ||| |:|||| :||
Db         71 EVLARHGSKIWVNEETKLVYFQGTKDTPLEHHLYVVSYEAAGEIVRLTTPGFSHSCSMSQ 130

Qy        148 HCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFE 207
              : | |:| ||:   | || :|||| |:||| |   |||:::::|   ||| ||||| |
Db        131 NFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWASMMEAASCPPDYVPPEIFHFH 190

Qy        208 STTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVV 267
              : :   ||||:|||| ||||||:|||||:|||||||||| |||:|| ||||||||| |
Db        191 TRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNTLASLGYAV 250

Qy        268 VVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGY 327
              |||| |||| |||:|||| | :|||:||:|||||::| :| |||| || |||||||:
Db        251 VVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAIHGWSYGGF 310

Qy        328 LSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              |||| |: :  :|:||||||||:|: ||||||
Db        311 LSLMGLIHKPQVFKVAIAGAPVTVWMAYDTGYT 343


RESULT 13
O75273
ID   O75273      PRELIMINARY;      PRT;   508 AA.
AC   O75273;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   R26984_1 (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA   Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA   Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA   Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA   Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA   Kronmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA   Kobayashi A., Olsen A.S., Carrano A.V.;
RT   "Sequence analysis of a 2.5 Mb region in 19p13.3.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005594; AAC33801.1; -.
DR   MEROPS; S09.019; -.
DR   InterPro; IPR000379; Ser_estrs_site.
FT   NON_TER       1       1
SQ   SEQUENCE   508 AA;  57750 MW;  2F4D6645BE2D2C89 CRC64;

  Query Match             48.8%;  Score 944;  DB 4;  Length 508;
  Best Local Similarity   54.7%;  Pred. No. 1.4e-77;
  Matches  185;  Conservative   48;  Mismatches   77;  Indels   28;  Gaps    5;

Qy         42 ILYEENDESEVEIIHVTSPMLETRRADSFRY-----PKTGTAN-PKVT-------FKMSE 88
              ::|||     : :  :  |  ::   |   :      |||  :  |||        :  ||
Db         75 VVYEEVTNVWINVHDIFYPFPQSEGEDELCFLRANECKTGFCHLYKVTAVLKSQGYDWSE 134

Qy         89 IMIDAEGR------IIVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHS 142
                 ||        | |:|  :||||:||||:||||||||||   ||: |||  |:|||
Db        135 PFSPGEGEQSLTNAIWVNEETKLVYFQGTKDTPLEHHLYVVSYEAAGEIVRLTTPGFSHS 194

Qy        143 CCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPE 202
              | :||: | |:| ||:   | || :|||| |:|||  |   |||:::::|         :
Db        195 CSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWASMMEAA---------K 245

Qy        203 IFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLAS 262
              || | : :   ||||:|||| ||||||:|||||:|||||||||| |||:|| ||||||||
Db        246 IFHFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNTLAS 305

Qy        263 LGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGW 322
              ||| ||||| |||| |||:|||| | :|||:||:|||||||::| :| |||| || ||||
Db        306 LGYAVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAIHGW 365

Qy        323 SYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              ||||:|||| |: :  :|:|||||||||:|: ||||||
Db        366 SYGGFLSLMGLIHKPQVFKVAIAGAPVTVWMAYDTGYT 403


RESULT 14
Q9HBM3
ID   Q9HBM3      PRELIMINARY;      PRT;   465 AA.
AC   Q9HBM3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Dipeptidyl peptidase 8 (Fragment).
GN   DPP8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=20467194; PubMed=11012666;
RA   Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
RA   Gorrell M.D.;
RT   "Cloning, expression and chromosomal localization of a novel human
RT   dipeptidyl peptidase (DPP) IV homolog, DPP8.";
RL   Eur. J. Biochem. 267:6140-6150(2000).
DR   EMBL; AF221636; AAG29768.1; -.
DR   MEROPS; S09.018; -.
FT   NON_TER       1       1
SQ   SEQUENCE   465 AA;  53197 MW;  22939EC0A4FE29CE CRC64;

  Query Match             44.5%;  Score 860.5;  DB 4;  Length 465;
  Best Local Similarity   40.5%;  Pred. No. 5.3e-70;
  Matches  186;  Conservative    0;  Mismatches    0;  Indels  273;  Gaps    2;

Qy         85 TGTANPKVTFKMSEIMIDAEGRII---------------------------------- 98
              ||||||||||||||||||||||||
Db          1 TGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSI 60

Qy         99 ------------------------------------------------------------ 98

Db         61 LLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIF 120
```

```
Qy         99 ------------------------------------------------------------ 98

Db        121 HVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIA 180

Qy         99 ----------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSH 141
                              ||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSH 240

Qy        142 SCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPP 201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPP 300

Qy        202 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLA 261
              ||||||||||||||||||||||||||||||||||||
Db        301 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQ---------------------- 340

Qy        262 SLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHG 321

Db        341 ------------------------------------------------------------ 340

Qy        322 WSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
                                ||||||||||||||||||
Db        341 --------------------VAIAGAPVTLWIFYDTGYT 359


RESULT 15
Q96NT8
ID   Q96NT8      PRELIMINARY;      PRT;   312 AA.
AC   Q96NT8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ30094.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
RA   Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
RA   Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA   Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA   Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K.,
RA   Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA   Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK054656; BAB70784.1; -.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00326; Peptidase_S9; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   312 AA;  35518 MW;  ABE940AFC5877717 CRC64;

  Query Match             41.3%;  Score 799;  DB 4;  Length 312;
  Best Local Similarity   66.2%;  Pred. No. 1.3e-64;
  Matches  143;  Conservative   31;  Mismatches   42;  Indels    0;  Gaps    0;

Qy        145 ISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIF 204
              :||: | |:| ||:   | || :|||| |:|||  |   |||:::::|   ||| |||||
Db          1 MSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWASMMEAASCPPDYVPPEIF 60

Qy        205 SFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLG 264
                | : :   ||||:|||| ||||||:|||||:|||||||||||| |||:|| |||||||||
Db         61 HFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNTLASLG 120

Qy        265 YVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSY 324
              | ||||| |||| |||:|||| | :|||:||:|||||||::| :| |||| || ||||||
Db        121 YAVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAIHGWSY 180

Qy        325 GGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360
              ||:|||| |: :  :|:|||| |||||:|: ||||||
Db        181 GGFLSLMGLIHKPQVFKVAIACAPVTVWMAYDTGYT 216


Search completed: October 15, 2003, 17:13:47
Job time : 25.0243 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:15:46 ; Search time 3930.37 Seconds
(without alignments)
3747.092 Million cell updates/sec

Title: US-10-070-464-7
Perfect score: 1933
Sequence: 1 EEDARSAGVATFVLQEEFDR.....RVAIAGAPVTLWIFYDTGYT 360

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10070464/runat_15102003_113553_24810/app_query.fasta_1.2652
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10070464_@CGN_1_1_13561_@runat_15102003_113553_24810 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1933	100.0	1083	9	AF221637	AF221637 Homo sapi
2	1880	97.3	4309	6	AX608737	AX608737 Sequence
3	1836.5	95.0	2649	6	AX354795	AX354795 Sequence
4	1836.5	95.0	2656	9	AY172659	AY172659 Homo sapi
5	1836.5	95.0	2671	6	AX608725	AX608725 Sequence
6	1836.5	95.0	3106	6	AX342633	AX342633 Sequence
7	1836.5	95.0	3127	9	AF221634	AF221634 Homo sapi
8	1836.5	95.0	3143	6	AX354793	AX354793 Sequence
9	1836.5	95.0	4535	9	BC040203	BC040203 Homo sapi
10	1816.5	94.0	4829	6	AX608735	AX608735 Sequence
11	1756.5	90.9	4799	10	BC043124	BC043124 Mus muscu
12	1534	79.4	3130	9	BC030688	BC030688 Homo sapi
13	1528	79.0	2161	6	BD157001	BD157001 Primer fo
14	1528	79.0	2161	9	AK027826	AK027826 Homo sapi
15	1522	78.7	2510	6	AX338497	AX338497 Sequence
16	1522	78.7	4685	6	AX608745	AX608745 Sequence
17	1521.5	78.7	4676	6	AX608743	AX608743 Sequence
18	1254.5	64.9	2668	6	AX405771	AX405771 Sequence
19	1254.5	64.9	2842	6	AX405770	AX405770 Sequence
20	1230	63.6	2778	9	AK000290	AK000290 Homo sapi
21	1229	63.6	4523	6	AX608731	AX608731 Sequence
22	1220.5	63.1	2602	9	AY172660	AY172660 Homo sapi
23	1220.5	63.1	2617	6	AX608727	AX608727 Sequence
24	1220.5	63.1	2676	6	AX524928	AX524928 Sequence
25	1220.5	63.1	3000	9	AF542510	AF542510 Homo sapi
26	1220.5	63.1	3243	9	HSM805448	AL834376 Homo sapi
27	1220.5	63.1	3716	6	AX480934	AX480934 Sequence
28	1220.5	63.1	4219	6	AX608751	AX608751 Sequence
29	1220.5	63.1	4232	9	BC037948	BC037948 Homo sapi
30	1220.5	63.1	4295	9	AF452102	AF452102 Homo sapi
31	1220.5	63.1	4302	6	AX608747	AX608747 Sequence
32	1201	62.1	4180	6	AX608759	AX608759 Sequence
33	1201	62.1	4263	6	AX608757	AX608757 Sequence
34	1132.5	58.6	4076	6	AX608755	AX608755 Sequence
35	1132.5	58.6	4159	6	AX608753	AX608753 Sequence
36	1113	57.6	4037	6	AX608763	AX608763 Sequence
37	1113	57.6	4120	6	AX608761	AX608761 Sequence
38	1089	56.3	753	6	AX524935	AX524935 Sequence
39	1086.5	56.2	1197	9	AF221635	AF221635 Homo sapi
40	984	50.9	2546	9	BC000970	BC000970 Homo sapi
41	975	50.4	2261	6	AX713363	AX713363 Sequence
42	975	50.4	2261	9	AK054656	AK054656 Homo sapi
43	860.5	44.5	1669	9	AF221636	AF221636 Homo sapi
44	735	38.0	1080	6	AX524932	AX524932 Sequence
45	718.5	37.2	3740	3	BT001499	BT001499 Drosophil

ALIGNMENTS

RESULT 1

```
AF221637
LOCUS       AF221637                1083 bp    mRNA    linear   PRI 05-NOV-2000
DEFINITION  Homo sapiens dipeptidyl peptidase 8 (DPP8) mRNA, partial cds,
            alternatively spliced.
ACCESSION   AF221637
VERSION     AF221637.1  GI:11095193
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1083)
  AUTHORS   Abbott,C.A., Yu,D.M., Woollatt,E., Sutherland,G.R., McCaughan,G.W.
            and Gorrell,M.D.
  TITLE     Cloning, expression and chromosomal localization of a novel human
            dipeptidyl peptidase (DPP) IV homolog, DPP8
  JOURNAL   Eur. J. Biochem. 267 (20), 6140-6150 (2000)
  MEDLINE   20467194
   PUBMED   11012666
REFERENCE   2  (bases 1 to 1083)
  AUTHORS   Abbott,C.A., Yu,D., McCaughan,G.W. and Gorrell,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JAN-2000) A.W. Morrow Gastroenterology and Liver
            Centre, Centenary Institute of Cell Biology and Cancer Medicine,
            Locked Bag No.6, Newtown, Sydney, NSW 2042, Australia
FEATURES             Location/Qualifiers
     source          1. .1083
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15q22"
                     /clone="stPBMCd3-3-1"
                     /cell_type="PHA stimulated PBMC"
     gene            <1. .>1083
                     /gene="DPP8"
     CDS             <1. .>1083
                     /gene="DPP8"
                     /note="serine protease; member of prolyl oligopetidase S9b
                     family belonging to enzyme clan SC; additional splice
                     variants of this mRNA are found in GenBank Accession
                     Numbers AF221634, AF221635 and AF221636"
                     /codon_start=2
                     /product="dipeptidyl peptidase 8"
                     /protein_id="AAG29769.1"
                     /db_xref="GI:11095194"
                     /translation="EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILY
                     EENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIVDEV
                     RRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQK
                     NPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGML
                     YKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSC
                     HRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALM
                     QRSDIFRVAIAGAPVTLWIFYDTGYT"
BASE COUNT      312 a    213 c    247 g    311 t
ORIGIN

Alignment Scores:
Pred. No.:              4.24e-191       Length:         1083
Score:                  1933.00         Matches:        360
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-10-070-464-7 (1-360) x AF221637 (1-1083)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 61

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 121

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 181

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 241

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIleValAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGTTGAT 301

Qy        101 GluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 GAAGTCAGAAGGCTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTG 361

Qy        121 TyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 TACGTAGTCAGTTACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCA 421

Qy        141 HisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 CATTCTTGCTGCATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAG 481

Qy        161 AsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLys 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 AATCCACACTGTGTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAA 541

Qy        181 ThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 ACAAAGGAATTTTGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCT 601

Qy        201 ProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLys 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 CCAGAAATTTTCTCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAG 661

Qy        221 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPro 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 CCTCATGATCTACAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCT 721

Qy        241 GlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 CAGGTGCAGTTGGTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTA 781

Qy        261 AlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 GCCTCTCTAGGTTATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTT 841

Qy        281 LysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGlu 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 AAATTTGAAGGCGCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAA 901

Qy        301 GlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHis 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 GGACTCCAATATCTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCAC 961

Qy        321 GlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePhe 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 GGCTGGTCCTATGGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTC 1021

Qy        341 ArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 AGGGTTGCTATTGCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 1081

RESULT 2
AX608737
LOCUS       AX608737                4309 bp    DNA     linear   PAT 17-FEB-2003
DEFINITION  Sequence 14 from Patent WO0231134.
ACCESSION   AX608737
VERSION     AX608737.1  GI:28404304
KEYWORDS    .
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Qi,S., Akinsanya,K.O., Riviere,P.J. and Junien,J.L.
  TITLE     Novel serine protease genes related to dppiv
  JOURNAL   Patent: WO 0231134-A 14 18-APR-2002;
            Ferring BV (NL)
FEATURES             Location/Qualifiers
     source          1. .4309
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT     1304 a    790 c    907 g   1308 t
ORIGIN

Alignment Scores:
Pred. No.:              7.46e-185       Length:         4309
Score:                  1880.00         Matches:        358
Percent Similarity:     98.90%          Conservative:   0
Best Local Similarity:  98.90%          Mismatches:     2
Query Match:            97.26%          Indels:         3
DB:                     6               Gaps:           0

US-10-070-464-7 (1-360) x AX608737 (1-4309)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1002

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1062

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1122

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1182

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIleValAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||      ||||||
Db       1183 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGATCCAA-GTTGAT 1241

Qy        101 GluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeu 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1242 GAAGTCAGAAGGCTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTG 1301

Qy        121 TyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSer 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1302 TACGTAGTCAGTTACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCA 1361

Qy        141 HisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLys 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1362 CATTCTTGCTGCATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAG 1421

Qy        161 AsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLys 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1422 AATCCACACTGTGTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAA 1481

Qy        181 ThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1482 ACAAAGGAATTTTGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCT 1541

Qy        201 ProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLys 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1542 CCAGAAATTTTCTCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAG 1601

Qy        221 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly--P 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
Db       1602 CCTCATGATCTACAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCTC 1661

Qy        240 roGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrL 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1662 CTCAGGTGCAGTTGGTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCC 1721

Qy        260 euAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCysHisArgGlyL 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1722 TAGCCTCTCTAGGTTATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGC 1781

Qy        280 euLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1782 TTAAATTTGAAGGCGCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGG 1841

Qy        300 luGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleH 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1842 AAGGACTCCAATATCTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCC 1901

Qy        320 isGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIleP 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1902 ACGGCTGGTCCTATGGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCT 1961

Qy        340 heArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1962 TCAGGGTTGCTATTGCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACA 2021

Qy        360 hr 360
              ||
Db       2022 CG 2023

RESULT 3
AX354795
LOCUS       AX354795               2649 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 3 from Patent WO0179473.
ACCESSION   AX354795
VERSION     AX354795.1  GI:18619528
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Meyers,R.A. and Williamson,M.
  TITLE     21953, a human prolyl oligopeptidase family member and uses thereof
  JOURNAL   Patent: WO 0179473-A 3 25-OCT-2001;
            Millennium Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .2649
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT     803 a    514 c    586 g    746 t
ORIGIN

Alignment Scores:
Pred. No.:              1.36e-180       Length:         2649
Score:                  1836.50         Matches:        360
Percent Similarity:     67.54%          Conservative:   0
Best Local Similarity:  67.54%          Mismatches:     0
Query Match:            95.01%          Indels:         173
DB:                     6               Gaps:           1

US-10-070-464-7 (1-360) x AX354795 (1-2649)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        730 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 789

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        790 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 849

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        850 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 909
Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
Db        910 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 969
Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
Db        970 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1029
Qy         98 ---------------------------------------------------------- 98
Db       1030 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1089
Qy         98 ---------------------------------------------------------- 98
Db       1090 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1149
Qy         98 ---------------------------------------------------------- 98
Db       1150 ACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1209
Qy         98 ---------------------------------------------------------- 98
Db       1210 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1269
Qy         98 ---------------------------------------------------------- 98
Db       1270 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1329
Qy         98 ---------------------------------------------------------- 98
Db       1330 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1389
Qy         98 ---------------------------------------------------------- 98
Db       1390 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1449
Qy         98 ---------------------------------------------------------- 98
Db       1450 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1509
Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
Db       1510 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1569
Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
Db       1570 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1629
Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
Db       1630 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1689
Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
Db       1690 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1749
Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
Db       1750 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 1809
Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
Db       1810 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 1869
Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
Db       1870 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 1929
Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
Db       1930 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 1989

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
Db       1990 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2049
Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
Db       2050 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2109
Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
Db       2110 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2169
Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
Db       2170 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2229
Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
Db       2230 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2289
Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
Db       2290 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2328
```

```
RESULT 4
AY172659
LOCUS       AY172659               2656 bp    mRNA    linear   PRI 08-JAN-2003
DEFINITION  Homo sapiens dipeptidyl peptidase IV-related protein-1 (DPRP1)
            mRNA, complete cds.
ACCESSION   AY172659
VERSION     AY172659.1  GI:27549549
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2656)
  AUTHORS   Qi,S., Akinsanya,K., Riviere,P. and Junien,J.-L.
  TITLE     Novel Serine Protease Genes Related To DPPIV
  JOURNAL   Patent: US (WO 0231134)-A 18-APR-2002;
REFERENCE   2  (bases 1 to 2656)
  AUTHORS   Qi,S., Akinsanya,K., Riviere,P. and Junien,J.-L.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-2002) Ferring Research Institute, 3550 General
            Atomics Ct., San Diego, CA 92121, USA
FEATURES             Location/Qualifiers
     source          1. .2656
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15q22.1-q22.2"
     gene            1. .2656
                     /gene="DPRP1"
     CDS             8. .2656
                     /gene="DPRP1"
                     /codon_start=1
                     /product="dipeptidyl peptidase IV-related protein-1"
                     /protein_id="AAO17261.1"
                     /db_xref="GI:27549550"
                     /translation="MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQ
                     LKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEI
                     PKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSG
                     TFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSN
                     DIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETT
                     PSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIMI
                     DAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLIS
                     PELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSHEEEIEF
                     IFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGR
                     HGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCD
                     FFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFES
                     TTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYV
```

```
                     VVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSY
                     GGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQ
                     AEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPES
                     GEHYELHLLHYLQENLGSRIAALKVI"
BASE COUNT      804 a    517 c    588 g    747 t
ORIGIN

Alignment Scores:
Pred. No.:              1.36e-180       Length:         2656
Score:                  1836.50         Matches:        360
Percent Similarity:     67.54%          Conservative:   0
Best Local Similarity:  67.54%          Mismatches:     0
Query Match:            95.01%          Indels:         173
DB:                     9               Gaps:           1

US-10-070-464-7 (1-360) x AY172659 (1-2656)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        737 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 796

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        797 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 856

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        857 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 916

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        917 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 976

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1036

Qy         98 ---------------------------------------------------------- 98

Db       1037 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1096

Qy         98 ----------------------------------------------------*----- 98

Db       1097 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1156

Qy         98 ---------------------------------------------------------- 98

Db       1157 ACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1216

Qy         98 ---------------------------------------------------------- 98

Db       1217 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1276

Qy         98 ---------------------------------------------------------- 98

Db       1277 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1336

Qy         98 ---------------------------------------------------------- 98

Db       1337 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1396

Qy         98 ---------------------------------------------------------- 98

Db       1397 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1456

Qy         98 ---------------------------------------------------------- 98

Db       1457 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1516

Qy         99 --------------------------------ValAspGluValArgArgLeuValTyr 107
                                              |||||||||||||||||||||||||
Db       1517 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1576

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1577 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1636

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1637 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1696

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1697 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1756

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1757 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 1816

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1817 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 1876

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1877 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 1936

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1937 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 1996

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1997 CGATTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2056

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2057 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2116

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2117 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2176

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2177 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2236

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2237 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2296

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||||
Db       2297 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2335

RESULT 5
AX608725
LOCUS       AX608725               2671 bp    DNA     linear   PAT 17-FEB-2003
DEFINITION  Sequence 2 from Patent WO0231134.
ACCESSION   AX608725
VERSION     AX608725.1  GI:28404298
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Qi,S., Akinsanya,K.O., Riviere,P.J. and Junien,J.L.
  TITLE     Novel serine protease genes related to dppiv
  JOURNAL   Patent: WO 0231134-A 2 18-APR-2002;
            Ferring BV (NL)
FEATURES             Location/Qualifiers
     source          1. .2671
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
```

```
BASE COUNT      805 a    524 c    594 g    748 t
ORIGIN

Alignment Scores:
Pred. No.:                1.37e-180       Length:          2671
Score:                    1836.50         Matches:         360
Percent Similarity:       67.54%          Conservative:    0
Best Local Similarity:    67.54%          Mismatches:      0
Query Match:              95.01%          Indels:          173
DB:                       6               Gaps:            1

US-10-070-464-7 (1-360) x AX608725 (1-2671)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        737 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 796

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        797 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 856

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        857 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 916

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        917 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 976

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1036

Qy         98 ---------------------------------------------------------- 98

Db       1037 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1096

Qy         98 ---------------------------------------------------------- 98

Db       1097 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1156

Qy         98 ---------------------------------------------------------- 98

Db       1157 ACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1216

Qy         98 ---------------------------------------------------------- 98

Db       1217 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1276

Qy         98 ---------------------------------------------------------- 98

Db       1277 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1336

Qy         98 ---------------------------------------------------------- 98

Db       1337 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1396

Qy         98 ---------------------------------------------------------- 98

Db       1397 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1456

Qy         98 ---------------------------------------------------------- 98

Db       1457 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1516

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                 |||||||||||||||||||||||||
Db       1517 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1576

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1577 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1636

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1637 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1696

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1697 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1756

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1757 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 1816

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1817 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 1876

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1877 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 1936

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1937 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 1996

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1997 CGATTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2056

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2057 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2116

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2117 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2176

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2177 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2236

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2237 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2296

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||||
Db       2297 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2335

RESULT 6
AX342633
LOCUS       AX342633               3106 bp    DNA     linear   PAT 12-JAN-2002
DEFINITION  Sequence 30 from Patent WO0198468.
ACCESSION   AX342633
VERSION     AX342633.1  GI:18152030
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Yue,H., Elliott,V.S., Gandhi,A.R., Lal,P., Au-Young,J.,
            Tribouley,C.M., Delegeane,A.M., Baughn,M.R., Nguyen,D.B., Lee,E.A.,
            Hafalia,A., Khan,F.A., Walia,N.K., Yao,M.G., Lu,D.A., Patterson,C.,
            Tang,Y.T., Walsh,R.T., Azimzai,Y., Ramkumar,J., Xu,Y. and Reddy,R.
  JOURNAL   Patent: WO 0198468-A 30 27-DEC-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .3106
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 7160544CB1"
BASE COUNT      928 a    633 c    704 g    841 t
```

```
ORIGIN

Alignment Scores:
Pred. No.:                1.65e-180       Length:          3106
Score:                    1836.50         Matches:         360
Percent Similarity:       67.54%          Conservative:    0
Best Local Similarity:    67.54%          Mismatches:      0
Query Match:              95.01%          Indels:          173
DB:                       6               Gaps:            1

US-10-070-464-7 (1-360) x AX342633 (1-3106)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        932 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 991

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        992 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1051

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1052 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1111

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1112 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1171

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1172 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1231

Qy         98 ---------------------------------------------------------- 98

Db       1232 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1291

Qy         98 ---------------------------------------------------------- 98

Db       1292 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1351

Qy         98 ---------------------------------------------------------- 98

Db       1352 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1411

Qy         98 ---------------------------------------------------------- 98

Db       1412 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1471

Qy         98 ---------------------------------------------------------- 98

Db       1472 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1531

Qy         98 ---------------------------------------------------------- 98

Db       1532 CACGAAGAGGAAATTGAGTTTATTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1591

Qy         98 ---------------------------------------------------------- 98

Db       1592 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1651

Qy         98 ---------------------------------------------------------- 98

Db       1652 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1711

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                 |||||||||||||||||||||||
Db       1712 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1771

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1772 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1831

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1832 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1891

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1892 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1951

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1952 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2011

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2012 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 2071

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2072 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 2131

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2132 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 2191

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2192 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2251

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2252 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2311

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2312 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2371

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2372 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2431

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2432 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2491

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||||
Db       2492 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2530

RESULT 7
AF221634
LOCUS       AF221634               3127 bp    mRNA    linear   PRI 05-NOV-2000
DEFINITION  Homo sapiens dipeptidyl peptidase 8 (DPP8) mRNA, complete cds.
ACCESSION   AF221634
VERSION     AF221634.1  GI:11095187
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3127)
  AUTHORS   Abbott,C.A., Yu,D.M., Woollatt,E., Sutherland,G.R., McCaughan,G.W.
            and Gorrell,M.D.
  TITLE     Cloning, expression and chromosomal localization of a novel human
            dipeptidyl peptidase (DPP) IV homolog, DPP8
  JOURNAL   Eur. J. Biochem. 267 (20), 6140-6150 (2000)
  MEDLINE   20467194
   PUBMED   11012666
REFERENCE   2  (bases 1 to 3127)
  AUTHORS   Abbott,C.A., Yu,D., McCaughan,G.W. and Gorrell,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JAN-2000) A.W. Morrow Gastroenterology and Liver
            Centre, Centenary Institute of Cell Biology and Cancer Medicine,
            Locked Bag No.6, Newtown, Sydney, NSW 2042, Australia
FEATURES             Location/Qualifiers
```

```
     source          1. .3127
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15q22"
     gene            1. .3127
                     /gene="DPP8"
     CDS             214. .2862
                     /gene="DPP8"
                     /note="serine protease; member of prolyl oligopetidase S9b
                     family belonging to enzyme clan SC; alternatively spliced
                     forms of this mRNA are found in AF221635, AF221636 and
                     AF221637"
                     /codon_start=1
                     /product="dipeptidyl peptidase 8"
                     /protein_id="AAG29766.1"
                     /db_xref="GI:11095188"
                     /translation="MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQ
                     LKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEI
                     PKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSG
                     TFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSN
                     DIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETT
                     PSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIMI
                     DAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLIS
                     PELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWININHDIFHVFPQSHEEEIEF
                     IFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGR
                     HGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCD
                     FFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFES
                     TTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYV
                     VVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSY
                     GGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQ
                     AEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPES
                     GEHYELHLLHYLQENLGSRIAALKVI"
BASE COUNT      943 a    637 c    706 g    841 t
ORIGIN

Alignment Scores:
Pred. No.:              1.67e-180       Length:         3127
Score:                  1836.50         Matches:        360
Percent Similarity:     67.54%          Conservative:   0
Best Local Similarity:  67.54%          Mismatches:     0
Query Match:            95.01%          Indels:         173
DB:                     9               Gaps:           1

US-10-070-464-7 (1-360) x AF221634 (1-3127)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1002

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1062

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1122

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1182

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1242

Qy         98 ---------------------------------------------------------- 98

Db       1243 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1302

Qy         98 ---------------------------------------------------------- 98

Db       1303 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1362

Qy         98 ---------------------------------------------------------- 98

Db       1363 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1422

Qy         98 ---------------------------------------------------------- 98

Db       1423 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1482

Qy         98 ---------------------------------------------------------- 98

Db       1483 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1542

Qy         98 ---------------------------------------------------------- 98

Db       1543 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1602

Qy         98 ---------------------------------------------------------- 98

Db       1603 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1662

Qy         98 ---------------------------------------------------------- 98

Db       1663 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1722

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                               |||||||||||||||||||||||||
Db       1723 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1782

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1842

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1902

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1962

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1963 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2022

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2023 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 2082

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2083 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 2142

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2143 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 2202

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2203 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2262

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2263 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2322

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2323 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2382

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2383 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2442
```

```
Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2443 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2502

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              |||||||||||||||||||||||||||||||||||||
Db       2503 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2541

RESULT 8
AX354793
LOCUS       AX354793               3143 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 1 from Patent WO0179473.
ACCESSION   AX354793
VERSION     AX354793.1  GI:18619526
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Meyers,R.A. and Williamson,M.
  TITLE     21953, a human prolyl oligopeptidase family member and uses thereof
  JOURNAL   Patent: WO 0179473-A 1 25-OCT-2001;
            Millennium Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .3143
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CDS             229. .2877
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD22990.1"
                     /db_xref="GI:18619527"
                     /translation="MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQ
                     LKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEI
                     PKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSG
                     TFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSN
                     DIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETT
                     PSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIMI
                     DAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLIS
                     PELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWININHDIFHVFPQSHEEEIEF
                     IFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGR
                     HGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCD
                     FFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFES
                     TTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYV
                     VVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSY
                     GGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQ
                     AEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPES
                     GEHYELHLLHYLQENLGSRIAALKVI"
BASE COUNT      943 a    644 c    712 g    844 t
ORIGIN

Alignment Scores:
Pred. No.:              1.68e-180       Length:         3143
Score:                  1836.50         Matches:        360
Percent Similarity:     67.54%          Conservative:   0
Best Local Similarity:  67.54%          Mismatches:     0
Query Match:            95.01%          Indels:         173
DB:                     6               Gaps:           1

US-10-070-464-7 (1-360) x AX354793 (1-3143)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1017

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1077

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1137

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1197

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1257

Qy         98 ---------------------------------------------------------- 98

Db       1258 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1317

Qy         98 ---------------------------------------------------------- 98

Db       1318 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1377

Qy         98 ---------------------------------------------------------- 98

Db       1378 ACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1437

Qy         98 ---------------------------------------------------------- 98

Db       1438 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1497

Qy         98 ---------------------------------------------------------- 98

Db       1498 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1557

Qy         98 ---------------------------------------------------------- 98

Db       1558 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1617

Qy         98 ---------------------------------------------------------- 98

Db       1618 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1677

Qy         98 ---------------------------------------------------------- 98

Db       1678 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1737

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                               |||||||||||||||||||||||||
Db       1738 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1797

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1798 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1857

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1858 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1917

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1918 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1977

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1978 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2037

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2038 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 2097

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2098 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 2157

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2158 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 2217

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2218 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2277

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2278 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2337

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2338 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2397

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2398 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2457

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2458 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2517

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              |||||||||||||||||||||||||||||||||||||||
Db       2518 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2556

RESULT 9
BC040203
LOCUS       BC040203               4535 bp    mRNA    linear   PRI 03-DEC-2002
DEFINITION  Homo sapiens, similar to dipeptidylpeptidase 8, clone MGC:42570
            IMAGE:4824813, mRNA, complete cds.
ACCESSION   BC040203
VERSION     BC040203.1  GI:26007916
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4535)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 71 Row: k Column: 23.
FEATURES             Location/Qualifiers
     source          1. .4535
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:42570 IMAGE:4824813"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
     CDS             1629. .4277
                     /codon_start=1
                     /product="similar to dipeptidylpeptidase 8"
                     /protein_id="AAH40203.1"
                     /db_xref="GI:26007917"
                     /db_xref="LocusID:54878"
                     /translation="MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQ
                     LKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEI
                     PKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSG
                     TFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHGN
                     DIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETT
                     PSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIMI
                     DAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLIS
                     PELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSHEEEIEF
                     IFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGR
                     HGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCD
                     FFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFES
                     TTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYV
                     VVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSY
                     GGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQ
                     AEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPES
                     GEHYELHLLHYLQENLGSRIAALKVI"
BASE COUNT     1313 a    918 c    946 g   1358 t
ORIGIN

Alignment Scores:
Pred. No.:              2.63e-180      Length:          4535
Score:                  1836.50        Matches:         360
Percent Similarity:     67.54%         Conservative:    0
Best Local Similarity:  67.54%         Mismatches:      0
Query Match:            95.01%         Indels:          173
DB:                     9              Gaps:            1

US-10-070-464-7 (1-360) x BC040203 (1-4535)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2358 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 2417

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2418 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 2477

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2478 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 2537

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2538 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 2597

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2598 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 2657

Qy         98 ---------------------------------------------------------- 98

Db       2658 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 2717

Qy         98 ---------------------------------------------------------- 98

Db       2718 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 2777

Qy         98 ---------------------------------------------------------- 98

Db       2778 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 2837

Qy         98 ---------------------------------------------------------- 98

Db       2838 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 2897

Qy         98 ----------
```

```
Db       2898 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 2957

Qy         98 ------------------------------------------------------------ 98

Db       2958 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 3017

Qy         98 ------------------------------------------------------------ 98

Db       3018 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 3077

Qy         98 ------------------------------------------------------------ 98

Db       3078 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 3137

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                 |||||||||||||||||||||||||
Db       3138 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 3197

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3198 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 3257

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3258 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 3317

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3318 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 3377

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3378 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 3437

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3438 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 3497

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3498 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 3557

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3558 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 3617

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3618 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 3677

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3678 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 3737

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3738 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 3797

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3798 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 3857

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3858 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 3917

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
               |||||||||||||||||||||||||||||||||||||
Db       3918 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 3956

RESULT 10
AX608735
```

```
LOCUS       AX608735                4829 bp    DNA     linear   PAT 17-FEB-2003
DEFINITION  Sequence 12 from Patent WO0231134.
ACCESSION   AX608735
VERSION     AX608735.1  GI:28404303
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Qi,S., Akinsanya,K.O., Riviere,P.J. and Junien,J.L.
  TITLE     Novel serine protease genes related to dppiv
  JOURNAL   Patent: WO 0231134-A 12 18-APR-2002;
            Ferring BV (NL)
FEATURES             Location/Qualifiers
     source          1. .4829
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT     1466 a    886 c   1017 g   1460 t
ORIGIN

Alignment Scores:
Pred. No.:              3.41e-178       Length:         4829
Score:                  1816.50         Matches:        360
Percent Similarity:     67.29%          Conservative:   0
Best Local Similarity:  67.29%          Mismatches:     0
Query Match:            93.97%          Indels:         175
DB:                     6               Gaps:           1

US-10-070-464-7 (1-360) x AX608735 (1-4829)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1002

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1062

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1122

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1182

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1242

Qy         98 ------------------------------------------------------------ 98

Db       1243 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1302

Qy         98 ------------------------------------------------------------ 98

Db       1303 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1362

Qy         98 ------------------------------------------------------------ 98

Db       1363 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1422

Qy         98 ------------------------------------------------------------ 98

Db       1423 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1482

Qy         98 ------------------------------------------------------------ 98

Db       1483 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1542

Qy         98 ------------------------------------------------------------ 98
```

```
Db       1543 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1602

Qy         98 ------------------------------------------------------------ 98

Db       1603 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1662

Qy         98 ------------------------------------------------------------ 98

Db       1663 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1722

Qy         99 --------------------------------ValAspGluValArgArgLeuValTyr 107
                                              ||||||||||||||||||||||||||
Db       1723 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1782

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1842

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1902

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1962

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1963 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2022

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2023 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 2082

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2083 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 2142

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGly--ProGlnValGlnLeuValAsnA 247
               ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db       2143 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCTCCTCAGGTGCAGTTGGTGAATA 2202

Qy        247 snArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValV 267
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2203 ATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGG 2262

Qy        267 alValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheL 287
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2263 TTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTA 2322

Qy        287 ysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaS 307
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2323 AATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTT 2382

Qy        307 erArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyT 327
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2383 CTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGAT 2442

Qy        327 yrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyA 347
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2443 ACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGG 2502

Qy        347 laProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
               ||||||||||||||||||||||||||||||||||||||||
Db       2503 CCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2543

RESULT 11
BC043124
LOCUS       BC043124                4799 bp    mRNA    linear   ROD 22-APR-2003
DEFINITION  Mus musculus dipeptidylpeptidase 8, mRNA (cDNA clone MGC:58057
            IMAGE:6410075), complete cds.
ACCESSION   BC043124
VERSION     BC043124.1  GI:27695449
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4799)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 4799)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            cDNA Library Preparation: M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
            Thomas L. Casavant.
            Web site: http://genome.uiowa.edu
            Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
            Bonaldo,M.F., Akabogu,I., Bair,T., Bair,J., Crouch,K., Davis,A.,
            Fishler,K., Keppel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K.,
            Scheetz,T., Smith,C., Snir,E., Tack,D., Trout,K., Walters,J.,
            Casavant,T., Soares,M.B.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series:  Plate:  Row:  Column: 0.
FEATURES                Location/Qualifiers
     source             1. .4799
                        /organism="Mus musculus"
                        /mol_type="mRNA"
                        /strain="C57BL/6"
                        /db_xref="taxon:10090"
                        /clone="MGC:58057 IMAGE:6410075"
                        /tissue_type="Brain, enriched mouse brain 12.5dp"
                        /clone_lib="NIH_BMAP_FO0"
                        /lab_host="DH10B"
                        /note="Vector: pYX-ASC"
     gene               1. .4799
                        /gene="Dpp8"
                        /db_xref="LocusID:74388"
                        /db_xref="MGI:1921638"
     CDS                288. .2966
                        /codon_start=1
                        /product="Dpp8 protein"
```

```
                     /protein_id="AAH43124.1"
                     /db_xref="GI:27695450"
                     /db_xref="LocusID:74388"
                     /translation="MKIPSGRCNMAAAMETEQLGVEIFETAECEEGNGESQDRPKLEP
                     FYVERYSWSQLKKLLADTRKYHGYMMAKAPHDFMFVKRTDPDGPHSDRVYYLAMSGEN
                     RENTLFYSEIPKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGI
                     AAYDYHPGSGTFLFQAGSGIYHIKDGGPHGFTQQPLRPNLVETSCPNIRMDPKLCPAD
                     PDWIAFIHSNDIWISNLVTREERRITYVHNELANMEEDPRSAGVATFVLQEEFDRYSG
                     YWWCPQAERTPSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPK
                     VTFKMSEIVVDAAGGIIDVIDKELVQPFEILFEGVEYIARAGWTPEGKHAWSILLDRS
                     QTHLQIVLISPELFIPVEDDAMDRQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVF
                     PQTHEDEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEITI
                     TSGEWEVLGRHGSNIWVDEARKLVYFEGTKDSPLEHHLYVTSYANPGEVVRLTDRGYS
                     HSCCLSRHCDFFISKYSNQKNPHCVSLYKLSSPEDDPVHKTKEFWATILDSAGPLPDY
                     TPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFR
                     LNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASQYDFIDL
                     DRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQ
                     GYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQ
                     ERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI"
BASE COUNT     1373 a   1015 c   1103 g   1308 t
ORIGIN

Alignment Scores:
Pred. No.:                5.81e-172       Length:          4799
Score:                    1756.50         Matches:         343
Percent Similarity:       65.67%          Conservative:    7
Best Local Similarity:    64.35%          Mismatches:      10
Query Match:              90.87%          Indels:          173
DB:                       10              Gaps:            1

US-10-070-464-7 (1-360) x BC043124 (1-4799)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1047 GAAGAGGATCCCAGATCAGCTGGTGTTGCCACCTTTGTTCTTCAAGAAGAATTTGACAGA 1106

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||:::||||||    ||||||||||||||||||||||||
Db       1107 TACTCTGGCTACTGGTGGTGTCCCCAAGCGGAAAGAACTCCTAGTGGTGGTAAAATTCTT 1166

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1167 AGAATTCTCTATGAAGAAAATGATGAATCTGAGGTGGAGATTATTCATGTTACGTCCCCC 1226

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1227 ATGTTGGAAACAAGGAGGGCAGATTCCTTTCGTTATCCCAAAACAGGCACAGCAAACCCA 1286

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||:::::||||||   |||   ||||||
Db       1287 AAGGTCACTTTCAAGATGTCGGAGATTGTTGTTGATGCTGCAGGAGGGATTATAGATGTC 1346

Qy         98 ---------------------------------------------------------- 98

Db       1347 ATAGATAAAGAACTGGTTCAACCTTTCGAGATTCTGTTTGAGGGAGTTGAATATATTGCC 1406

Qy         98 ---------------------------------------------------------- 98

Db       1407 AGAGCCGGATGGACTCCAGAGGGAAAACATGCCTGGTCCATCCTACTAGACCGTTCCCAG 1466

Qy         98 ---------------------------------------------------------- 98

Db       1467 ACTCACCTGCAGATAGTTCTGATCTCCCCTGAGTTATTCATCCCAGTAGAAGATGATGCC 1526

Qy         98 ---------------------------------------------------------- 98

Db       1527 ATGGACAGACAGAGACTTATAGAGTCAGTTCCTGACTCTGTGACACCACTGATCATCTAT 1586

Qy         98 ---------------------------------------------------------- 98

Db       1587 GAAGAAACAACAGACATCTGGATAAATATCCACGATATTTTTCATGTTTTTCCTCAAACT 1646

Qy         98 ---------------------------------------------------------- 98

Db       1647 CATGAAGATGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTTCGTCATCTG 1706

Qy         98 ---------------------------------------------------------- 98

Db       1707 TATAAAATCACATCCATTTTAAAGGAGAGCAAATATAAACGGTCCAGTGGTGGACTACCT 1766

Qy         98 ---------------------------------------------------------- 98

Db       1767 GCCCCAAGTGATTTCAAGTGTCCTATCAAAGAAGAAATAACAATTACCAGTGGTGAATGG 1826

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                 |||||||||   |||:::|||||||||
Db       1827 GAAGTACTTGGCCGGCATGGATCTAATATCTGGGTTGATGAAGCCAGAAAGCTGGTGTAC 1886

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||   ||||||   |||
Db       1887 TTTGAAGGCACCAAAGACTCTCCTTTGGAGCATCACCTGTACGTGACCAGTTATGCAAAC 1946

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||   |||||||||||||||||||||||||||||||:::|||:::
Db       1947 CCTGGAGAAGTGGTGAGGCTGACTGACCGTGGCTACTCACACTCCTGCTGCCTCAGCCGG 2006

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2007 CATTGTGACTTCTTCATAAGTAAGTACAGCAACCAGAAGAATCCACACTGTGTGTCCCTC 2066

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||      ||||||||||||||||||||||||
Db       2067 TACAAACTCTCAAGTCCTGAGGATGACCCAGTTCATAAAACAAAGGAATTTTGGGCCACC 2126

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2127 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTACACCCCTCCAGAAATTTTTTCTTTTGAA 2186

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2187 AGTACTACTGGATTTACACTGTATGGAATGTTGTATAAGCCTCATGACCTACAACCTGGA 2246

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2247 AAGAAATACCCCACTGTGTTATTCATATATGGTGGTCCCCAGGTGCAGCTGGTGAACAAT 2306

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2307 CGGTTTAAAGGAGTCAAGTATTTCCGCCTGAACACCCTGGCCTCCCTGGGTTATGTGGTT 2366

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2367 GTGGTGATAGACAACAGGGGATCCTGTCACCGAGGACTTAAATTTGAAGGCGCCTTTAAA 2426

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2427 TATAAAATGGGTCAAATAGAAATCGATGATCAAGTGGAAGGACTCCAGTACCTAGCATCT 2486

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2487 CAGTATGACTTCATTGACTTGGATCGAGTGGGCATCCACGGCTGGTCCTATGGTGGCTAC 2546

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2547 CTCTCCCTGATGGCATTAATGCAGAGGTCGGATATCTTCCGGGTGGCTATTGCTGGGGCC 2606

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||
Db       2607 CCAGTCACCCTGTGGATCTTCTATGATACAGGATACACG 2645

RESULT 12
BC030688
LOCUS       BC030688               3130 bp    mRNA    linear   PRI 30-MAY-2002
DEFINITION  Homo sapiens, Similar to dipeptidylpeptidase 8, clone MGC:26191
            IMAGE:4822550, mRNA, complete cds.
ACCESSION   BC030688
```

```
VERSION     BC030688.1  GI:21265132
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3130)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 33 Row: d Column: 5
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 18450277.
FEATURES             Location/Qualifiers
     source          1. .3130
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:26191 IMAGE:4822550"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
     CDS             372. .2867
                     /codon_start=1
                     /product="Similar to dipeptidylpeptidase 8"
                     /protein_id="AAH30688.1"
                     /db_xref="GI:21265133"
                     /translation="MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQ
                     LKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEI
                     PKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSG
                     TFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSN
                     DIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETT
                     PSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIMI
                     DAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLIS
                     PELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSHEEEIEF
                     IFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGR
                     HGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCD
                     FFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFES
                     TTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYV
                     VVVIDNRGSCHRGLKFEGAFKYKMVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYY
                     LGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERH
                     SIRVPESGEHYELHLLHYLQENLGSRIAALKVI"
BASE COUNT      955 a    626 c    717 g    832 t
ORIGIN

Alignment Scores:
Pred. No.:                4.55e-149       Length:          3130
Score:                    1534.00         Matches:         309
Percent Similarity:       57.97%          Conservative:    0
Best Local Similarity:    57.97%          Mismatches:      0
Query Match:              79.36%          Indels:          224
DB:                       9               Gaps:            2

US-10-070-464-7 (1-360) x BC030688 (1-3130)
```

```
Qy       1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
Db    1101 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1160

Qy      21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
Db    1161 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1220

Qy      41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
Db    1221 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1280

Qy      61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
Db    1281 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1340

Qy      81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
Db    1341 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1400

Qy      98 ---------------------------------------------------------- 98
Db    1401 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1460

Qy      98 ---------------------------------------------------------- 98
Db    1461 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1520

Qy      98 ---------------------------------------------------------- 98
Db    1521 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1580

Qy      98 ---------------------------------------------------------- 98
Db    1581 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1640

Qy      98 ---------------------------------------------------------- 98
Db    1641 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1700

Qy      98 ---------------------------------------------------------- 98
Db    1701 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1760

Qy      98 ---------------------------------------------------------- 98
Db    1761 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1820

Qy      98 ---------------------------------------------------------- 98
Db    1821 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1880

Qy      99 ------------------------------ValAspGluValArgArgLeuValTyr 107
Db    1881 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1940

Qy     108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
Db    1941 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 2000

Qy     128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
Db    2001 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 2060

Qy     148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
Db    2061 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 2120

Qy     168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
Db    2121 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2180
```

```
Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2181 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 2240

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2241 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 2300

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2301 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 2360

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2361 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2420

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2421 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2480

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              |||||||||
Db       2481 TATAAAATG---------------------------------------------------- 2489

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327

Db       2489 ---------------------------------------------------------- 2489

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
                                                  ||||||||||||||||||
Db       2490 ------------------------------------------GTTGCTATTGCTGGGGCC 2507

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||
Db       2508 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2546
```

```
RESULT 13
BD157001
LOCUS       BD157001                2161 bp    DNA     linear   PAT 17-JAN-2003
DEFINITION  Primer for synthesizing full-length cDNA and use thereof.
ACCESSION   BD157001
VERSION     BD157001.1  GI:27862759
KEYWORDS    JP 2002191363-A/11844.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2161)
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
            Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE     Primer for synthesizing full-length cDNA and use thereof
  JOURNAL   Patent: JP 2002191363-A 11844 09-JUL-2002;
            HELIX RESEARCH INSTITUTE
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002191363-A/11844
            PD   09-JUL-2002
            PF   28-JUL-2000 JP 2000280990
            PI   TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
            PI   SAITO,
            PI   JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
            PI   KEIICHI NAGAI,TETSUJI OTSUKI
            PC
            C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
             10,
            PC   C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
             Primer for synthesizing full-length cDNA and use.thereof FH   Key
                    Location/Qualifiers
            FT   CDS             (16). .(1911).
FEATURES             Location/Qualifiers
     source          1. .2161
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
```

```
BASE COUNT      662 a    429 c    452 g    618 t
ORIGIN

Alignment Scores:
Pred. No.:                1.21e-148       Length:          2161
Score:                    1528.00         Matches:         308
Percent Similarity:       57.79%          Conservative:    0
Best Local Similarity:    57.79%          Mismatches:      1
Query Match:              79.05%          Indels:          224
DB:                       6               Gaps:            2

US-10-070-464-7 (1-360) x BD157001 (1-2161)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        148 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 207

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 267

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        268 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 327

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        328 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 387

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        388 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 447

Qy         98 ---------------------------------------------------------- 98

Db        448 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 507

Qy         98 ---------------------------------------------------------- 98

Db        508 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 567

Qy         98 ---------------------------------------------------------- 98

Db        568 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 627

Qy         98 ---------------------------------------------------------- 98

Db        628 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 687

Qy         98 ---------------------------------------------------------- 98

Db        688 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 747

Qy         98 ---------------------------------------------------------- 98

Db        748 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 807

Qy         98 ---------------------------------------------------------- 98

Db        808 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 867

Qy         98 ---------------------------------------------------------- 98

Db        868 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 927

Qy         99 ----A----------------------------ValAspGluValArgArgLeuValTyr 107
                                               ||||||||||||||||||||||||
Db        928 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 987

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        988 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1047
```

```
Qy       128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
             ||||||||||||||||||||||||||||||      ||||||||||||||||||||||||||
Db      1048 CCTGGAGAGGTGACAAGGCTGACTGACCGTAGCTACTCACATTCTTGCTGCATCAGTCAG 1107

Qy       148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1108 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1167

Qy       168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1168 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 1227

Qy       188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1228 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 1287

Qy       208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1288 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 1347

Qy       228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1348 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 1407

Qy       248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1408 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 1467

Qy       268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1468 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 1527

Qy       288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
             |||||||||
Db      1528 TATAAAATG--------------------------------------------------- 1536

Qy       308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327

Db      1536 ------------------------------------------------------------ 1536

Qy       328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
                                                         ||||||:||||||||||||
Db      1537 -----------------------------------------GTTGCTATTGCTGGGGCC 1554

Qy       348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
             |||||||||||||||||||||||||||||||||||||||
Db      1555 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 1593
```

```
RESULT 14
AK027826
LOCUS       AK027826                2161 bp    mRNA    linear   PRI 01-AUG-2002
DEFINITION  Homo sapiens cDNA FLJ14920 fis, clone PLACE1007416, weakly similar
            to DIPEPTIDYL PEPTIDASE IV (EC 3.4.14.5).
ACCESSION   AK027826
VERSION     AK027826.1  GI:14042789
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2161)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology; cDNA library construction,
            5'- & 3'-end one pass sequencing and clone selection: Helix
            Research Institute (supported by Japan Key Technology Center etc.)
            and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1..2161
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="PLACE1007416"
                     /tissue_type="placenta"
                     /clone_lib="PLACE1"
                     /note="cloning vector: pME18SFL3"
     CDS             16..1914
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB55395.1"
                     /db_xref="GI:14042790"
                     /translation="MDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNELANM
                     EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVT
                     SPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGV
                     EYIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSV
                     TPLIIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKY
                     KRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLE
                     HHLYVVSYVNPGEVTRLTDRSYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPED
                     DPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTV
                     LFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMV
                     AIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFL
                     DENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQENLGS
                     RIAALKVI"
BASE COUNT     662 a    429 c    452 g    618 t
ORIGIN

Alignment Scores:
Pred. No.:              1.21e-148      Length:          2161
Score:                  1528.00        Matches:         308
Percent Similarity:     57.79%         Conservative:    0
Best Local Similarity:  57.79%         Mismatches:      1
Query Match:            79.05%         Indels:          224
DB:                     9              Gaps:            2

US-10-070-464-7 (1-360) x AK027826 (1-2161)

Qy         1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       148 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 207

Qy        21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       208 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 267

Qy        41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       268 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 327

Qy        61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       328 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 387

Qy        81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       388 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 447

Qy        98 ------------------------------------------------------------ 98

Db       448 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 507
```

```
Qy        98 ------------------------------------------------------------ 98

Db       508 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 567

Qy        98 ------------------------------------------------------------ 98

Db       568 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 627

Qy        98 ------------------------------------------------------------ 98

Db       628 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 687

Qy        98 ------------------------------------------------------------ 98

Db       688 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 747

Qy        98 ------------------------------------------------------------ 98

Db       748 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 807

Qy        98 ------------------------------------------------------------ 98

Db       808 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 867

Qy        98 ------------------------------------------------------------ 98

Db       868 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 927

Qy        99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                 |||||||||||||||||||||||||
Db       928 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 987

Qy       108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       988 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1047

Qy       128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
             ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db      1048 CCTGGAGAGGTGACAAGGCTGACTGACCGTAGCTACTCACATTCTTGCTGCATCAGTCAG 1107

Qy       148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1108 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1167

Qy       168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1168 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 1227

Qy       188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1228 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 1287

Qy       208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1288 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 1347

Qy       228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1348 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 1407

Qy       248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1408 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 1467

Qy       268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1468 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 1527

Qy       288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
             |||||||||
Db      1528 TATAAAATG------------------------------------------------- 1536

Qy       308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327

Db      1536 ------------------------------------------------------------ 1536

Qy       328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
                                                   ||||||||||||||||||
Db      1537 ----------------------------------------GTTGCTATTGCTGGGGCC 1554

Qy       348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
             ||||||||||||||||||||||||||||||||||||||||
Db      1555 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 1593

RESULT 15
AX338497
LOCUS       AX338497               2510 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 16 from Patent WO0183775.
ACCESSION   AX338497
VERSION     AX338497.1  GI:18128919
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Delegeane,A.M., Lal,P., Hafalia,A., Patterson,C., Walia,N.K.,
            Kearney,L., Tribouley,C.M., Khan,F.A., Yao,M.G., Baughn,M.R.,
            Azimzai,Y., Elliott,V.S., Nguyen,D.B., Gandhi,A.R., Yang,J.,
            Hernandez,R., Policky,J.L., Lu,D.A., Reddy,R. and Tang,Y.T.
  TITLE     Proteases
  JOURNAL   Patent: WO 0183775-A 16 08-NOV-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .2510
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 376067CB1"
BASE COUNT      777 a    494 c    527 g    712 t
ORIGIN

Alignment Scores:
Pred. No.:              6.12e-148       Length:         2510
Score:                  1522.00         Matches:        309
Percent Similarity:     58.08%          Conservative:   0
Best Local Similarity:  58.08%          Mismatches:     0
Query Match:            78.74%          Indels:         223
DB:                     6               Gaps:           2

US-10-070-464-7 (1-360) x AX338497 (1-2510)

Qy         1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       609 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 668

Qy        21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       669 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 728

Qy        41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       729 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 788

Qy        61 MetLeuGluThrArg-ArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPr 80
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       789 ATGTTGGAAACAAGGCAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCC 848

Qy        80 oLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       849 TAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGT 908

Qy        98 ------------------------------------------------------------ 98

Db       909 CATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGC 968
```

```
Qy        98 ------------------------------------------------------------ 98
Db       969 CAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCA 1028
Qy        98 ------------------------------------------------------------ 98
Db      1029 GACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGT 1088
Qy        98 ------------------------------------------------------------ 98
Db      1089 TATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTA 1148
Qy        98 ------------------------------------------------------------ 98
Db      1149 TGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAG 1208
Qy        98 ------------------------------------------------------------ 98
Db      1209 TCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTT 1268
Qy        98 ------------------------------------------------------------ 98
Db      1269 ATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCC 1328
Qy        98 ------------------------------------------------------------ 98
Db      1329 TGCTCCAACTGTCACTTGGATGATCACATTCATGAGATCTCTAGGAACTCCATCCTGTAT 1388
Qy        99 ----------------------------ValAspGluValArgArgLeuValTyrPheGl 109
                                         ||||||||||||||||||||||||||||||||
Db      1389 GTGTGTGACACATATAGTTGAGATCCAAGTTGATGAAGTCAGAAGGCTGGTATATTTTGA 1448
Qy       109 uGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsnProGl 129
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1449 AGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAATCCTGG 1508
Qy       129 yGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGlnHisCy 149
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1509 AGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAGCACTG 1568
Qy       149 sAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeuTyrLy 169
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1569 TGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTTTACAA 1628
Qy       169 sLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThrIleLe 189
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1629 GCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACCATTTT 1688
Qy       189 uAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGluSerTh 209
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1689 GGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAAAGTAC 1748
Qy       209 rThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGlyLysLy 229
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1749 TACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGAAAGAA 1808
Qy       229 sTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsnArgPh 249
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1809 ATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAATCGGTT 1868
Qy       249 eLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValValValVa 269
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1869 TAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTTGTAGT 1928
Qy       269 lIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLysTyrLy 289
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1929 GATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAATATAA 1988
Qy       289 sMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTy 309
             ||||
Db      1989 AATG---------------------------------------------------------- 1992
Qy       309 rAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSe 329

Db      1992 ---------------------------------------------------------- 1992
Qy       329 rLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProVa 349
                                                 ||||||||||||||||||||||
Db      1993 ----------------------------------GTTGCTATTGCTGGGGCCCCAGT 2015
Qy       349 lThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
                 |||||||||||||||||||||||||||||||
Db      2016 CACTCTGTGGATCTTCTATGATACAGGATACACG 2049
```

Search completed: October 16, 2003, 00:05:27
Job time : 3960.37 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:14:56 ; Search time 264.512 Seconds
(without alignments)
3673.932 Million cell updates/sec

Title: US-10-070-464-7
Perfect score: 1933
Sequence: 1 EEDARSAGVATFVLQEEFDR.....RVAIAGAPVTLWIFYDTGYT 360

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10070464/runat_15102003_113552_24799/app_query.fasta_1.2652
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10070464_@CGN_1_1_0_@runat_15102003_113552_24799 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1933	100.0	1083	22	AAC85697	Nucleotide sequenc
2	1880	97.3	4309	24	ABK83328	cDNA encoding huma
3	1836.5	95.0	2643	24	AAH99935	Coding sequence of
4	1836.5	95.0	2649	25	ABX12255	cDNA encoding huma
5	1836.5	95.0	2671	24	ABK83322	cDNA encoding huma
6	1836.5	95.0	3106	24	ABK12892	Human protease PRT
7	1836.5	95.0	3120	22	AAC85694	Nucleotide sequenc
8	1836.5	95.0	3120	24	AAD38956	Human dipeptidyl p
9	1836.5	95.0	3143	24	AAH99934	cDNA encoding 2195
10	1816.5	94.0	4829	24	ABK83327	cDNA encoding huma
11	1528	79.0	2161	22	AAH15009	Human cDNA sequenc
12	1522	78.7	2510	24	AAD23843	Human protease PRT
13	1522	78.7	4685	24	ABK83332	cDNA encoding huma
14	1521.5	78.7	4676	24	ABK83331	cDNA encoding huma
15	1429	73.9	1821	24	ABV76411	Dipeptidyl peptida
16	1254.5	64.9	2668	24	ABN59775	Novel human coding
17	1254.5	64.9	2842	24	ABN59774	Novel human coding
18	1229	63.6	4523	24	ABK83325	cDNA encoding huma
19	1220.5	63.1	2495	24	AAD38957	Human dipeptidyl p
20	1220.5	63.1	2617	24	ABK83323	cDNA encoding huma
21	1220.5	63.1	3024	24	AAD38954	Human dipeptidyl p
22	1220.5	63.1	3716	24	ABQ75955	Human PMMM encodin
23	1220.5	63.1	4219	24	ABK83335	cDNA encoding huma
24	1220.5	63.1	4302	24	ABK83333	cDNA encoding huma
25	1204.5	62.3	2751	24	AAD38311	Murine dipeptidyl
26	1204.5	62.3	3287	24	AAD38955	Alternative versio
27	1201	62.1	4180	24	ABK83339	cDNA encoding huma
28	1201	62.1	4263	24	ABK83338	cDNA encoding huma
29	1183.5	61.2	3262	22	AAI57880	Human polynucleoti
30	1169.5	60.5	2801	22	AAI57896	Human polynucleoti
31	1169.5	60.5	2982	22	AAI59666	Human polynucleoti
32	1132.5	58.6	4076	24	ABK83337	cDNA encoding huma
33	1132.5	58.6	4159	24	ABK83336	cDNA encoding huma
34	1113	57.6	4037	24	ABK83341	cDNA encoding huma
35	1113	57.6	4120	24	ABK83340	cDNA encoding huma
36	1086.5	56.2	1197	22	AAC85695	Nucleotide sequenc
37	984	50.9	2952	24	ABK69090	DNA encoding human
38	980	50.7	3047	24	ABK69113	DNA encoding human
39	944	48.8	2027	21	AAC77137	Human ORFX ORF2692
40	860.5	44.5	1669	22	AAC85696	Nucleotide sequenc
41	722	37.4	3713	23	ABL10425	Drosophila melanog
42	722	37.4	3783	23	ABL06641	Drosophila melanog
43	654.5	33.9	2461	21	AAC75835	Human ORFX ORF1390
44	558	28.9	6225	23	ABL10424	Drosophila melanog
45	558	28.9	6228	23	ABL06640	Drosophila melanog

ALIGNMENTS

RESULT 1
AAC85697
ID AAC85697 standard; cDNA; 1083 BP.
XX
AC AAC85697;
XX
DT 29-JUN-2001 (first entry)
XX
DE Nucleotide sequence encoding human DPP8 244Glu-341Ile+515Val-776Thr.
XX
KW Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
KW dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;
KW growth hormone deficiency; glucose level; mucosal regeneration;
KW non-insulin dependent diabetes mellitus;glucose intolerance;
KW immunosuppression; ss.
XX
OS Homo sapiens.

```
XX
FH   Key             Location/Qualifiers
FT   CDS             2..1081
FT                   /*tag= a
FT                   /product= "Human DPP8 244Glu-341Ile+515Val-776Thr"
FT                   /note= "No stop codon given"
XX
PN   WO200119866-A1.
XX
PD   22-MAR-2001.
XX
PF   11-SEP-2000; 2000WO-AU01085.
XX
PR   10-SEP-1999;   99AU-0002762.
PR   18-FEB-2000; 2000AU-0005709.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Abbott CA,  Gorell MD;
XX
DR   WPI; 2001-281520/29.
DR   P-PSDB; AAB47190.
XX
PT   New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
PT   substrates, identifying inhibitors of DPP8 catalytic activity which
PT   have therapeutic uses, and for detecting activated T cells  -
XX
PS   Claim 21; Page 75-76; 78pp; English.
XX
CC   The sequences given in AAC85695-97 encode fragments of human
CC   dipeptidyl aminopeptidase (DPP8).  DPP8 has substrate specificity for
CC   H-Ala-Pro-pNA, H-Gly-Pro-pNA and H-Arg-Pro-pNA. Therefore, it is a
CC   prolyl oligopeptidase and a dipeptidyl peptidase, because it is capable
CC   of hydrolysing  the peptide bond C-terminal to Pro in each of these
CC   compounds.  DPP8 is homologous with human DPPIV.  DPP8 is useful for
CC   cleaving a substrate, and for detecting an activated T cell which
CC   involves measuring the level of DPP8 gene expression in a T cell. The
CC   level of DPP8 expression is detected by detecting the amount of DPP8
CC   RNA in the cell. It is also useful for identifying a molecule capable
CC   of inhibiting the cleavage of the substrate by DPP8.  Molecules
CC   identified as inhibiting DPP8 catalytic activity may be useful for
CC   treating diarrhoea, growth hormone deficiency, lowering glucose levels
CC   in non-insulin dependent diabetes mellitus and other disorders
CC   involving glucose intolerance, enhancing mucosal regeneration and
CC   as immunosuppressants.
XX
SQ   Sequence 1083 BP; 312 A; 213 C; 247 G; 311 T; 0 other;

Alignment Scores:
Pred. No.:              1.46e-216       Length:         1083
Score:                  1933.00         Matches:        360
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     22              Gaps:           0

US-10-070-464-7 (1-360) x AAC85697 (1-1083)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 61

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 121

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 181

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 241

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIleValAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGTTGAT 301

Qy        101 GluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 GAAGTCAGAAGGCTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTG 361

Qy        121 TyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 TACGTAGTCAGTTACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCA 421

Qy        141 HisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 CATTCTTGCTGCATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAG 481

Qy        161 AsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLys 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 AATCCACACTGTGTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAA 541

Qy        181 ThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 ACAAAGGAATTTTGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCT 601

Qy        201 ProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLys 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 CCAGAAATTTTCTCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAG 661

Qy        221 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPro 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 CCTCATGATCTACAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCT 721

Qy        241 GlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 CAGGTGCAGTTGGTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTA 781

Qy        261 AlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 GCCTCTCTAGGTTATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTT 841

Qy        281 LysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGlu 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 AAATTTGAAGGCGCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAA 901

Qy        301 GlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHis 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 GGACTCCAATATCTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCAC 961

Qy        321 GlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePhe 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 GGCTGGTCCTATGGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTC 1021

Qy        341 ArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 AGGGTTGCTATTGCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 1081

RESULT 2
ABK83328
ID   ABK83328 standard; cDNA; 4309 BP.
XX
AC   ABK83328;
XX
DT   12-AUG-2002  (first entry)
XX
DE   cDNA encoding human DPRP-1 splice variant #4.
XX
KW   Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW   DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW   diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW   heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW   ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW   dyskinesia; reproductive disorder; inflammatory disorder;
```

KW metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI; 2002-444178/47.
DR P-PSDB; ABG61597.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT -
XX
PS Disclosure; Page 67-68; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABK83322-ABK83343 encode human DPRP proteins.
XX
SQ Sequence 4309 BP; 1304 A; 790 C; 907 G; 1308 T; 0 other;

Alignment Scores:

Pred. No.:	1.95e-209	Length:	4309
Score:	1880.00	Matches:	358
Percent Similarity:	98.90%	Conservative:	0
Best Local Similarity:	98.90%	Mismatches:	2
Query Match:	97.26%	Indels:	3
DB:	24	Gaps:	0

US-10-070-464-7 (1-360) x ABK83328 (1-4309)

```
Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1002

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1062

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1122

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1182

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIleValAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGATCCAA-GTTGAT 1241

Qy        101 GluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1242 GAAGTCAGAAGGCTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTG 1301

Qy        121 TyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1302 TACGTAGTCAGTTACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCA 1361

Qy        141 HisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1362 CATTCTTGCTGCATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAG 1421

Qy        161 AsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLys 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1422 AATCCACACTGTGTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAA 1481

Qy        181 ThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1482 ACAAAGGAATTTTGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCT 1541

Qy        201 ProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLys 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1542 CCAGAAATTTTCTCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAG 1601

Qy        221 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly--P 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||  |  |
Db       1602 CCTCATGATCTACAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCTC 1661

Qy        240 roGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrL 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1662 CTCAGGTGCAGTTGGTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCC 1721

Qy        260 euAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCysHisArgGlyL 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1722 TAGCCTCTCTAGGTTATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGC 1781

Qy        280 euLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1782 TTAAATTTGAAGGCGCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGG 1841

Qy        300 luGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleH 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1842 AAGGACTCCAATATCTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCC 1901

Qy        320 isGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIleP 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1902 ACGGCTGGTCCTATGGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCT 1961

Qy        340 heArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1962 TCAGGGTTGCTATTGCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACA 2021

Qy        360 hr 360
              ||
Db       2022 CG 2023
```

RESULT 3
AAH99935
ID AAH99935 standard; cDNA; 2643 BP.
XX
AC AAH99935;
XX
DT 12-APR-2002 (first entry)
XX
DE Coding sequence of 21953 human prolyl oligopeptidase.
XX
KW 21953 prolyl oligopeptidase; antibody; proline; endopeptidase;
KW cancer; cardiovascular disease; autoimmune disease; atopic allergy;
KW neuronal disorder; vascular disorder; prostate disorder; cytostatic;
KW antidiabetic; antiarthritic; antiasthmatic; antiinflammatory;
KW diabetes mellitus; arthritis; multiple sclerosis; asthma;
KW Grave's disease; neuronal disorder; demyelinating disease; ss.
XX

OS Homo sapiens.
XX
PN WO200179473-A2.
XX
PD 25-OCT-2001.
XX
PF 11-APR-2001; 2001WO-US40483.
XX
PR 18-APR-2000; 2000US-197508P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Meyers RA, Williamson M;
XX
DR WPI; 2002-034353/04.
DR P-PSDB; AAG78415.
XX
PT New polypeptides 21953, member of human prolyl oligopeptidase family,
PT useful as diagnostic targets and therapeutic agents for controlling
PT cancer, lymphoma and leukemia -
XX
PS Claim 2; Page-; 121pp; English.
XX
CC This invention relates to an isolated 21953 human prolyl
CC oligopeptidase. Which is cytostatic, antidiabetic, antiarthritic,
CC neuroprotective, antithyroid, dermatological, antipsoriatic,
CC antiasthmatic, ophthalmological, antiinflammatory, nootropic,
CC antiparkinsonian, anticonvulsant, gynaecological, vasotropic,
CC antianginal, cardiant, antiatherosclerotic, anorectic and
CC metabolic in its action. Uses include gene therapy, expression or
CC activity of 21953 protein modulator, it is useful for identifying a
CC compound which binds to it and can be used in preventing, treating
CC or detecting a cellular proliferative or differentiative disorder.
CC The 21953 molecules can act as novel diagnostic targets and therapeutic
CC agents for controlling disorders associated with the aberrant activity
CC or degradation of peptide hormones e.g., disorders associated with cell
CC differentiation and proliferation such as cancer, immune function,
CC reproductive, neurological and cardiovascular function. The 21953
CC molecules are thus useful for treating and preventing cellular
CC proliferative and differentiative disorders, haematopoietic neoplastic
CC disorders, immune disorders such as autoimmune diseases, diabetes
CC mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
CC neuronal disorders, demyelinating diseases, vascular disorders and
CC metabolism or pain disorders. This sequence represents the cDNA
CC encoding sequence of 21953 human prolyl oligopeptidase. This
CC sequence represents the coding sequence of 21953, being the sequence
CC in between the start and the stop codon of the sequence represented in
CC AAH99934.
XX
SQ Sequence 2643 BP; 800 A; 514 C; 585 G; 744 T; 0 other;

Alignment Scores:

Pred. No.:	1.19e-204	Length:	2643
Score:	1836.50	Matches:	360
Percent Similarity:	67.54%	Conservative:	0
Best Local Similarity:	67.54%	Mismatches:	0
Query Match:	95.01%	Indels:	173
DB:	24	Gaps:	1

US-10-070-464-7 (1-360) x AAH99935 (1-2643)

```
Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        727 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 786

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        787 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 846

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        847 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 906

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        907 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 966

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        967 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1026

Qy         98 ---------------------------------------------------------- 98

Db       1027 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1086

Qy         98 ---------------------------------------------------------- 98

Db       1087 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1146

Qy         98 ---------------------------------------------------------- 98

Db       1147 ACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1206

Qy         98 ---------------------------------------------------------- 98

Db       1207 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1266

Qy         98 ---------------------------------------------------------- 98

Db       1267 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1326

Qy         98 ---------------------------------------------------------- 98

Db       1327 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1386

Qy         98 ---------------------------------------------------------- 98

Db       1387 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1446

Qy         98 ---------------------------------------------------------- 98

Db       1447 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1506

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                 |||||||||||||||||||||||||
Db       1507 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1566

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1567 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1626

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1627 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1686

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1687 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1746

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1747 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 1806

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1807 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 1866

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1867 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 1926

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1927 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 1986

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1987 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2046

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2047 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2106

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2107 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2166

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2167 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2226

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2227 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2286

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
                ||||||||||||||||||||||||||||||||||||
Db       2287 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2325

RESULT 4
ABX12255
ID   ABX12255 standard; cDNA; 2649 BP.
XX
AC   ABX12255;
XX
DT   19-MAY-2003  (first entry)
XX
DE   cDNA encoding human serine protease HIPHUM46.
XX
KW   Human; ss; gene; HIPHUM46; serine protease; gene therapy; osteoarthritis;
KW   serine protease activity modulation; dipeptidyl peptidase activity;
KW   musculoskeletal disease; Hepatitis B virus infection; myotonic dystrophy;
KW   Alzheimer's disease; parasupranuclear palsy; Huntington's disease;
KW   amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;
KW   irritable bowel syndrome; type I diabetes; faecal incontinence; anaemia;
KW   haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;
KW   colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;
KW   multiple sclerosis; chromosome 15q21-q22.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2649
FT                   /*tag=  a
FT                   /product= "HIPHUM46"
FT                   /note= "Serine protease"
XX
PN   GB2374869-A.
XX
PD   30-OCT-2002.
XX
PF   22-JAN-2002; 2002GB-0001404.
XX
PR   23-JAN-2001; 2001GB-0001760.
XX
PA   (GLAX ) GLAXO GROUP LTD.
XX
PI   Edbrooke MR,  Lewis AP;
XX
DR   WPI; 2003-150703/15.
DR   P-PSDB; ABU07720.
XX
PT   Identifying modulators of serine protease activity useful for treating
PT   musculoskeletal diseases, by contacting cell expressing a novel serine
PT   protease polypeptide with a compound and monitoring serine protease
PT   activity  -
XX
PS   Claim 12; Page 22-26; 38pp; English.
XX
CC   The invention relates to a method of identifying a substance that
CC   modulates serine protease activity, comprising contacting a cell such as
CC   a neuronal cell, lung cell, intestinal cell or a cell infected with a
CC   virus, expressing a serine protease polypeptide (HIPHUM 46), or its
CC   variant having dipeptidyl peptidase activity, or a serine protease
CC   isolated from the cell with a test substance and monitoring for serine
CC   protease activity. The method is useful for identifying a substance that
CC   modulates serine protease activity. A modulator of the serine protease is
CC   useful in the manufacture of a medicament for treatment or prophylaxis of
CC   a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus
CC   infection, Alzheimer's disease, parasupranuclear palsy, myotonic
CC   dystrophy, Huntington's disease or amyotrophic lateral sclerosis.
CC   Additional disease that may be treated using modulators of the serine
CC   protease include malabsorption syndromes, irritable bowel syndrome, lung
CC   disease, type I diabetes, faecal incontinence, haemorrhoids, proctitis,
CC   rectal polyps, small bowel tumours, colorectal tumours, anaemia,
CC   dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple
CC   sclerosis. The present sequence represents cDNA encoding the human serine
CC   protease HIPHUM46 which is located on chromosome 15q21-q22.
XX
SQ   Sequence 2649 BP; 803 A; 514 C; 585 G; 747 T; 0 other;

Alignment Scores:
Pred. No.:                1.19e-204      Length:          2649
Score:                    1836.50        Matches:         360
Percent Similarity:       67.54%         Conservative:    0
Best Local Similarity:    67.54%         Mismatches:      0
Query Match:              95.01%         Indels:          173
DB:                       25             Gaps:            1

US-10-070-464-7 (1-360) x ABX12255 (1-2649)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        730 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 789

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        790 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 849

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        850 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 909

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        910 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 969

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        970 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1029

Qy         98 ---------------------------------------------------------- 98

Db       1030 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1089

Qy         98 ---------------------------------------------------------- 98

Db       1090 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1149

Qy         98 ---------------------------------------------------------- 98

Db       1150 ACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1209

Qy         98 ---------------------------------------------------------- 98

Db       1210 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1269

Qy         98 ---------------------------------------------------------- 98

Db       1270 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1329

Qy         98 ---------------------------------------------------------- 98
```

```
Db       1330 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1389

Qy         98 ------------------------------------------------------------ 98

Db       1390 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1449

Qy         98 ------------------------------------------------------------ 98

Db       1450 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1509

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                 |||||||||||||||||||||||||
Db       1510 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1569

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1570 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1629

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1630 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1689

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1690 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1749

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1750 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 1809

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1810 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 1869

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1870 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 1929

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1930 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 1989

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1990 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2049

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2050 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2109

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2110 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2169

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2170 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2229

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2230 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2289

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
               |||||||||||||||||||||||||||||||||||||
Db       2290 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2328

RESULT 5
ABK83322
ID   ABK83322 standard; cDNA; 2671 BP.
XX
AC   ABK83322;
XX
DT   12-AUG-2002  (first entry)
XX
DE   cDNA encoding human DPPIV related serine protease DPRP-1.
XX
KW   Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW   DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW   diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW   heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW   ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW   dyskinesia; reproductive disorder; inflammatory disorder;
KW   metabolic disorder; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200239134-A2.
XX
PD   18-APR-2002.
XX
PF   12-OCT-2001; 2001WO-US31874.
XX
PR   12-OCT-2000; 2000US-240117P.
XX
PA   (FERR ) FERRING BV.
XX
PI   Qi S,  Akinsanya KO,  Riviere PJ,  Junien J;
XX
DR   WPI; 2002-444178/47.
DR   P-PSDB; ABG61591.
XX
PT   New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT   the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT   viral infections, cancers, allergies, neurological disorders, or pain
PT   -
XX
PS   Claim 1; Page 53-54; 113pp; English.
XX
CC   The present invention relates to the isolation of novel human serine
CC   proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC   proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC   and nucleic acids encoding them are useful for treating infections
CC   such as fungal, bacterial, protozoan and viral infections, particularly
CC   infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC   pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC   bulimia, Parkinson's disease, acute heart failure, hypotension,
CC   hypertension, urinary retention, osteoporosis, angina pectoris,
CC   stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC   psychotic and neurological disorders (e.g. anxiety, dementia, or
CC   schizophrenia), and dyskinesias. These may also be used in discovering
CC   therapeutic agents for the treatment of reproductive, inflammatory and
CC   metabolic disorders. ABK83322-ABK83343 encode human DPRP proteins.
XX
SQ   Sequence 2671 BP; 805 A; 524 C; 594 G; 748 T; 0 other;

Alignment Scores:
Pred. No.:                1.21e-204      Length:          2671
Score:                    1836.50        Matches:         360
Percent Similarity:       67.54%         Conservative:    0
Best Local Similarity:    67.54%         Mismatches:      0
Query Match:              95.01%         Indels:          173
DB:                       24             Gaps:            1

US-10-070-464-7 (1-360) x ABK83322 (1-2671)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        737 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 796

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        797 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 856

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        857 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 916

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        917 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 976

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1036

Qy         98 ---------------------------------------------------------- 98

Db       1037 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1096

Qy         98 ---------------------------------------------------------- 98

Db       1097 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1156

Qy         98 ---------------------------------------------------------- 98

Db       1157 ACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1216

Qy         98 ---------------------------------------------------------- 98

Db       1217 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1276

Qy         98 ---------------------------------------------------------- 98

Db       1277 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1336

Qy         98 ---------------------------------------------------------- 98

Db       1337 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1396

Qy         98 ---------------------------------------------------------- 98

Db       1397 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1456

Qy         98 ---------------------------------------------------------- 98

Db       1457 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1516

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                 |||||||||||||||||||||||
Db       1517 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1576

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1577 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1636

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1637 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1696

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1697 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1756

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1757 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 1816

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1817 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 1876

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1877 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 1936

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1937 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 1996
```

```
Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1997 CGATTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2056

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2057 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2116

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2117 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2176

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2177 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2236

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2237 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2296

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              |||||||||||||||||||||||||||||||||||||
Db       2297 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2335
```

RESULT 6
ABK12892
ID ABK12892 standard; cDNA; 3106 BP.
XX
AC ABK12892;
XX
DT 09-APR-2002 (first entry)
XX
DE Human protease PRTS-9 cDNA sequence.
XX
KW Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KW cell proliferative disorder; developmental disorder; epilepsy;
KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KW reproductive disorder; endometriosis; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 203..2851
FT /*tag= a
FT /product= "Human protease PRTS-9"
XX
PN WO200198468-A2.
XX
PD 27-DEC-2001.
XX
PF 13-JUN-2001; 2001WO-US19178.
XX
PR 16-JUN-2000; 2000US-212336P.
PR 22-JUN-2000; 2000US-213955P.
PR 29-JUN-2000; 2000US-215396P.
PR 07-JUL-2000; 2000US-216821P.
PR 14-JUL-2000; 2000US-218946P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM;
PI Delegeane AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
PI Walia NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT;
PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
PI Kallick DA;
XX
DR WPI; 2002-090437/12.
DR P-PSDB; AAU74749.
XX
PT Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful

```
PT   in the diagnosis, treatment and prevention of gastrointestinal (e.g.
PT   gastritis), cardiovascular (e.g. atherosclerosis) and cell
PT   proliferative (e.g. cancer)disorders -
XX
PS   Claim 5; Page 166-167; 177pp; English.
XX
CC   The present invention relates to twenty one new human proteases,
CC   referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
CC   polypeptides of the invention are useful in the diagnosis, treatment and
CC   prevention of gastrointestinal e.g. gastritis,, esophageal carcinoma and
CC   Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
CC   myocardial infarction, autoimmune/inflammatory e.g. acquired
CC   immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
CC   proliferative e.g. cancer, developmental e.g. Duchenne and Becker
CC   muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
CC   epilepsy and Alzheimer's disease and reproductive e.g. infertility and
CC   endometriosis disorders. Numerous other examples of each disorder are
CC   given in the specification. The present nucleic acid sequence encodes
CC   the human protease PRTS-9 protein of the invention.
XX
SQ   Sequence 3106 BP; 928 A; 633 C; 704 G; 841 T; 0 other;

Alignment Scores:
Pred. No.:                1.52e-204      Length:          3106
Score:                    1836.50        Matches:         360
Percent Similarity:       67.54%         Conservative:    0
Best Local Similarity:    67.54%         Mismatches:      0
Query Match:              95.01%         Indels:          173
DB:                       24             Gaps:            1

US-10-070-464-7 (1-360) x ABK12892 (1-3106)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        932 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 991

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        992 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1051

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1052 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1111

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1112 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1171

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1172 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1231

Qy         98 ---------------------------------------------------------- 98

Db       1232 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1291

Qy         98 ---------------------------------------------------------- 98

Db       1292 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1351

Qy         98 ---------------------------------------------------------- 98

Db       1352 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1411

Qy         98 ---------------------------------------------------------- 98

Db       1412 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1471

Qy         98 ---------------------------------------------------------- 98

Db       1472 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1531

Qy         98 ---------------------------------------------------------- 98

Db       1532 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1591

Qy         98 ---------------------------------------------------------- 98

Db       1592 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1651

Qy         98 ---------------------------------------------------------- 98

Db       1652 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1711

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                 ||||||||||||||||||||||||
Db       1712 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1771

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1772 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1831

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1832 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1891

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1892 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1951

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1952 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2011

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2012 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 2071

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2072 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 2131

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2132 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 2191

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2192 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2251

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2252 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2311

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2312 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2371

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2372 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2431

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2432 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2491

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||||
Db       2492 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2530

RESULT 7
AAC85694
ID   AAC85694 standard; cDNA; 3120 BP.
XX
AC   AAC85694;
XX
```

```
DT   29-JUN-2001  (first entry)
XX
DE   Nucleotide sequence of human DPP8.
XX
KW   Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
KW   dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;
KW   growth hormone deficiency; glucose level; mucosal regeneration;
KW   non-insulin dependent diabetes mellitus;glucose intolerance;
KW   immunosuppression; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             214..2862
FT                   /*tag=  a
FT                   /product= "Human DPP8"
XX
PN   WO200119866-A1.
XX
PD   22-MAR-2001.
XX
PF   11-SEP-2000; 2000WO-AU01085.
XX
PR   10-SEP-1999;   99AU-0002762.
PR   18-FEB-2000; 2000AU-0005709.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Abbott CA,  Gorell MD;
XX
DR   WPI; 2001-281520/29.
DR   P-PSDB; AAB47187.
XX
PT   New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
PT   substrates, identifying inhibitors of DPP8 catalytic activity which
PT   have therapeutic uses, and for detecting activated T cells  -
XX
PS   Claim 16; Fig 2; 78pp; English.
XX
CC   This sequence encodes human dipeptidyl aminopeptidase (DPP8).  DPP8
CC   has substrate specificity for H-Ala-Pro-pNA, H-Gly-Pro-pNA and
CC   H-Arg-Pro-pNA. Therefore, it is a prolyl oligopeptidase and a
CC   dipeptidyl peptidase, because it is capable of hydrolysing  the
CC   peptide bond C-terminal to Pro in each of these compounds.  DPP8
CC   is homologous with human DPPIV.  DPP8 is useful for cleaving a
CC   substrate, and for detecting an activated T cell which involves
CC   measuring the level of DPP8 gene expression in a T cell. The level
CC   of DPP8 expression is detected by detecting the amount of DPP8 RNA
CC   in the cell. It is also useful for identifying a molecule capable
CC   of inhibiting the cleavage of the substrate by DPP8.  Molecules
CC   identified as inhibiting DPP8 catalytic activity may be useful for
CC   treating diarrhoea, growth hormone deficiency, lowering glucose levels
CC   in non-insulin dependent diabetes mellitus and other disorders
CC   involving glucose intolerance, enhancing mucosal regeneration and
CC   as immunosuppressants.
XX
SQ   Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;

Alignment Scores:
Pred. No.:              1.53e-204       Length:         3120
Score:                  1836.50         Matches:        360
Percent Similarity:     67.54%          Conservative:   0
Best Local Similarity:  67.54%          Mismatches:     0
Query Match:            95.01%          Indels:         173
DB:                     22              Gaps:           1

US-10-070-464-7 (1-360) x AAC85694 (1-3120)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1002

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1062

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1122

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1182

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1242

Qy         98 ---------------------------------------------------------- 98

Db       1243 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1302

Qy         98 ---------------------------------------------------------- 98

Db       1303 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1362

Qy         98 ---------------------------------------------------------- 98

Db       1363 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1422

Qy         98 ---------------------------------------------------------- 98

Db       1423 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1482

Qy         98 ---------------------------------------------------------- 98

Db       1483 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1542

Qy         98 ---------------------------------------------------------- 98

Db       1543 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1602

Qy         98 ---------------------------------------------------------- 98

Db       1603 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1662

Qy         98 ---------------------------------------------------------- 98

Db       1663 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1722

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                 ||||||||||||||||||||||||
Db       1723 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1782

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1842

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1902

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1962

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1963 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2022

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2023 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 2082

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspGluLeuProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2083 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 2142

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2143 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 2202

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2203 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2262

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2263 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2322

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2323 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2382

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2383 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2442

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2443 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2502

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||
Db       2503 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2541

RESULT 8
AAD38956
ID   AAD38956 standard; cDNA; 3120 BP.
XX
AC   AAD38956;
XX
DT   23-SEP-2002  (first entry)
XX
DE   Human dipeptidyl peptidase 8 (DPP8) cDNA.
XX
KW   Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
KW   autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;
KW   graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
KW   antiviral; enzyme; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             214..2862
FT                   /*tag=  a
FT                   /product= "Human DPP8 protein"
XX
PN   WO200234900-A1.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-AU01388.
XX
PR   27-OCT-2000; 2000AU-0001078.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Abbott CA,  Gorrell MD;
XX
DR   WPI; 2002-454646/48.
DR   P-PSDB; AAE24170.
XX
PT   New dipeptidyl peptidase (DPP) peptides, useful for screening
PT   inhibitors of DPP catalytic activity, which may be employed to treat
PT   e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PT   rejection and HIV infection -
XX
PS   Example; Fig 1; 91pp; English.
XX
CC   The present invention relates to dipeptidyl peptidase (DPP) proteins and
CC   polynucleotides encoding such proteins. The DPP peptides are useful for
CC   screening inhibitors of DPP catalytic activity. The inhibitors are useful
CC   for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
CC   rejection and HIV (human immuno deficiency virus) infection. The present
CC   sequence is human DPP8 cDNA.
XX
SQ   Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;

Alignment Scores:
Pred. No.:                1.53e-204      Length:          3120
Score:                    1836.50        Matches:         360
Percent Similarity:       67.54%         Conservative:    0
Best Local Similarity:    67.54%         Mismatches:      0
Query Match:              95.01%         Indels:          173
DB:                       24             Gaps:            1

US-10-070-464-7 (1-360) x AAD38956 (1-3120)

Qy           1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         943 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1002

Qy          21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1003 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1062

Qy          41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1063 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1122

Qy          61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1123 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1182

Qy          81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1183 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1242

Qy          98 ---------------------------------------------------------- 98

Db        1243 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1302

Qy          98 ---------------------------------------------------------- 98

Db        1303 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1362

Qy          98 ---------------------------------------------------------- 98

Db        1363 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1422

Qy          98 ---------------------------------------------------------- 98

Db        1423 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1482

Qy          98 ---------------------------------------------------------- 98

Db        1483 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1542

Qy          98 ---------------------------------------------------------- 98

Db        1543 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1602

Qy          98 ---------------------------------------------------------- 98

Db        1603 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1662

Qy          98 ---------------------------------------------------------- 98

Db        1663 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1722

Qy          99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
```

```
                                                    ||||||||||||||||||||||||||
Db       1723 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1782

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1842

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1902

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1962

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1963 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2022

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2023 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 2082

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2083 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 2142

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2143 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 2202

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2203 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2262

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2263 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2322

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2323 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2382

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2383 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2442

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2443 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2502

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||||
Db       2503 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2541

RESULT 9
AAH99934
ID   AAH99934 standard; cDNA; 3143 BP.
XX
AC   AAH99934;
XX
DT   12-APR-2002  (first entry)
XX
DE   cDNA encoding 21953 human prolyl oligopeptidase.
XX
KW   21953 prolyl oligopeptidase; human; proline; endopeptidase;
KW   cancer; cardiovascular disease; autoimmune disease; atopic allergy;
KW   neuronal disorder; vascular disorder; prostate disorder; cytostatic;
KW   antidiabetic; antiarthritic; antiasthmatic; antiinflammatory;
KW   diabetes mellitus; arthritis; multiple sclerosis; asthma;
KW   Grave's disease; neuronal disorder; demyelinating disease; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             229..2877
FT                   /*tag=  a
FT                   /product= "21953 prolyl oligopeptidase"
FT                   /note= "This region is specifically claimed in
FT                   claim 2"
XX
PN   WO200179473-A2.
XX
PD   25-OCT-2001.
XX
PF   11-APR-2001; 2001WO-US40483.
XX
PR   18-APR-2000; 2000US-197508P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Meyers RA,  Williamson M;
XX
DR   WPI; 2002-034353/04.
DR   P-PSDB; AAG78415.
XX
PT   New polypeptides 21953, member of human prolyl oligopeptidase family,
PT   useful as diagnostic targets and therapeutic agents for controlling
PT   cancer, lymphoma and leukemia  -
XX
PS   Claim 7; Page 100-102; 121pp; English.
XX
CC   This invention relates to an isolated 21953 human prolyl
CC   oligopeptidase. Which is cytostatic, antidiabetic, antiarthritic,
CC   neuroprotective, antithyroid, dermatological, antipsoriatic,
CC   antiasthmatic, ophthalmological, antiinflammatory, nootropic,
CC   antiparkinsonian, anticonvulsant, gynaecological, vasotropic,
CC   antianginal, cardiant, antiatherosclerotic, anorectic and
CC   metabolic in its action. Uses include gene therapy, expression or
CC   activity of 21953 protein modulator, it is useful for identifying a
CC   compound which binds to it and can be used in preventing, treating
CC   or detecting a cellular proliferative or differentiative disorder.
CC   The 21953 molecules can act as novel diagnostic targets and therapeutic
CC   agents for controlling disorders associated with the aberrant activity
CC   or degradation of peptide hormones e.g., disorders associated with cell
CC   differentiation and proliferation such as cancer, immune function,
CC   reproductive, neurological and cardiovascular function. The 21953
CC   molecules are thus useful for treating and preventing cellular
CC   proliferative and differentiative disorders, haematopoietic neoplastic
CC   disorders, immune disorders such as autoimmune diseases, diabetes
CC   mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
CC   neuronal disorders, demyelinating diseases, vascular disorders and
CC   metabolism or pain disorders. This sequence represents the cDNA
CC   encoding sequence of 21953 human prolyl oligopeptidase.
XX
SQ   Sequence 3143 BP; 943 A; 644 C; 712 G; 844 T; 0 other;

Alignment Scores:
Pred. No.:                1.54e-204      Length:          3143
Score:                    1836.50        Matches:         360
Percent Similarity:       67.54%         Conservative:    0
Best Local Similarity:    67.54%         Mismatches:      0
Query Match:              95.01%         Indels:          173
DB:                       24             Gaps:            1

US-10-070-464-7 (1-360) x AAH99934 (1-3143)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1017

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1077

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
```

```
Db       1078 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1137

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro  80

Db       1138 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1197

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------  98

Db       1198 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1257

Qy         98 ------------------------------------------------------------  98

Db       1258 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1317

Qy         98 ------------------------------------------------------------  98

Db       1318 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1377

Qy         98 ------------------------------------------------------------  98

Db       1378 ACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1437

Qy         98 ------------------------------------------------------------  98

Db       1438 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1497

Qy         98 ------------------------------------------------------------  98

Db       1498 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1557

Qy         98 ------------------------------------------------------------  98

Db       1558 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1617

Qy         98 ------------------------------------------------------------  98

Db       1618 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1677

Qy         98 ------------------------------------------------------------  98

Db       1678 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1737

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107

Db       1738 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1797

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127

Db       1798 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1857

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147

Db       1858 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1917

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167

Db       1918 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1977

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187

Db       1978 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2037

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207

Db       2038 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 2097

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227

Db       2098 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 2157

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247

Db       2158 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 2217

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267

Db       2218 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2277

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287

Db       2278 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2337

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307

Db       2338 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2397

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327

Db       2398 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2457

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347

Db       2458 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2517

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360

Db       2518 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2556
```

RESULT 10
ABK83327
ID ABK83327 standard; cDNA; 4829 BP.
XX
AC ABK83327;
XX
DT 12-AUG-2002 (first entry)
XX
DE cDNA encoding human DPRP-1 splice variant #3.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI; 2002-444178/47.
DR P-PSDB; ABG61596.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT -
XX
PS Disclosure; Page 65-66; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly

```
CC   infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC   pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC   bulimia, Parkinson's disease, acute heart failure, hypotension,
CC   hypertension, urinary retention, osteoporosis, angina pectoris,
CC   stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC   psychotic and neurological disorders (e.g. anxiety, dementia, or
CC   schizophrenia), and dyskinesias. These may also be used in discovering
CC   therapeutic agents for the treatment of reproductive, inflammatory and
CC   metabolic disorders. ABK83322-ABK83343 encode human DPRP proteins.
XX
SQ   Sequence 4829 BP; 1466 A; 886 C; 1017 G; 1460 T; 0 other;

Alignment Scores:
Pred. No.:              6.56e-202       Length:         4829
Score:                  1816.50         Matches:        360
Percent Similarity:     67.29%          Conservative:   0
Best Local Similarity:  67.29%          Mismatches:     0
Query Match:            93.97%          Indels:         175
DB:                     24              Gaps:           1

US-10-070-464-7 (1-360) x ABK83327 (1-4829)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1002

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1062

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1122

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1182

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1242

Qy         98 ---------------------------------------------------------- 98

Db       1243 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1302

Qy         98 ---------------------------------------------------------- 98

Db       1303 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1362

Qy         98 ---------------------------------------------------------- 98

Db       1363 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1422

Qy         98 ---------------------------------------------------------- 98

Db       1423 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1482

Qy         98 ---------------------------------------------------------- 98

Db       1483 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1542

Qy         98 ---------------------------------------------------------- 98

Db       1543 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1602

Qy         98 ---------------------------------------------------------- 98

Db       1603 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1662

Qy         98 ---------------------------------------------------------- 98

Db       1663 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1722

Qy         99 ----------------------------------ValAspGluValArgArgLeuValTyr 107
                                                |||||||||||||||||||||||||
Db       1723 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1782

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1842

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1902

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1962

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1963 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2022

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2023 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 2082

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2083 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 2142

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGly--ProGlnValGlnLeuValAsnA 247
              ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db       2143 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCTCCTCAGGTGCAGTTGGTGAATA 2202

Qy        247 snArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValV 267
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2203 ATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGG 2262

Qy        267 alValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheL 287
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2263 TTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTA 2322

Qy        287 ysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaS 307
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2323 AATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTT 2382

Qy        307 erArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyT 327
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2383 CTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGAT 2442

Qy        327 yrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyA 347
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2443 ACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGG 2502

Qy        347 laProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              |||||||||||||||||||||||||||||||||||||||
Db       2503 CCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2543

RESULT 11
AAH15009
ID   AAH15009 standard; cDNA; 2161 BP.
XX
AC   AAH15009;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:12963.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
```

PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 12963; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2161 BP; 662 A; 429 C; 452 G; 618 T; 0 other;

Alignment Scores:

Pred. No.:	1.38e-168	Length:	2161
Score:	1528.00	Matches:	308
Percent Similarity:	57.79%	Conservative:	0
Best Local Similarity:	57.79%	Mismatches:	1
Query Match:	79.05%	Indels:	224
DB:	22	Gaps:	2

US-10-070-464-7 (1-360) x AAH15009 (1-2161)

```
Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        148 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 207

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 267

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        268 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 327

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        328 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 387

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        388 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 447

Qy         98 ---------------------------------------------------------- 98

Db        448 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 507

Qy         98 ---------------------------------------------------------- 98

Db        508 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 567

Qy         98 ---------------------------------------------------------- 98

Db        568 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 627

Qy         98 ---------------------------------------------------------- 98

Db        628 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 687

Qy         98 ---------------------------------------------------------- 98

Db        688 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 747

Qy         98 ---------------------------------------------------------- 98

Db        748 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 807

Qy         98 ---------------------------------------------------------- 98

Db        808 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 867

Qy         98 ---------------------------------------------------------- 98

Db        868 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 927

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                 ||||||||||||||||||||||||
Db        928 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 987

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        988 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1047

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1048 CCTGGAGAGGTGACAAGGCTGACTGACCGTAGCTACTCACATTCTTGCTGCATCAGTCAG 1107

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1108 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1167

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1168 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 1227

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1228 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 1287

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1288 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 1347

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1348 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 1407

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1408 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 1467
```

```
Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1468 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 1527

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              |||||||||
Db       1528 TATAAAATG---------------------------------------------------- 1536

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327

Db       1536 ---------------------------------------------------------- 1536

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
                                                  ||||||||||||||||||
Db       1537 -----------------------------------------GTTGCTATTGCTGGGGCC 1554

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              |||||||||||||||||||||||||||||||||||||
Db       1555 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 1593
```

RESULT 12
AAD23843
ID AAD23843 standard; cDNA; 2510 BP.
XX
AC AAD23843;
XX
DT 07-MAR-2002 (first entry)
XX
DE Human protease PRTS-2 cDNA.
XX
KW Human; protease; PRTS-2; tranquilliser; gene therapy; vaccine; allergy;
KW infection; dermatitis; arteriosclerosis; rheumatoid arthritis; hepatitis;
KW atherosclerosis; psoriasis; Alzheimer's disease; mental disorder; cancer;
KW gastrointestinal disorder; Cushing's syndrome; seizure; glaucoma; stroke;
KW epithelial disorder; urticaria; anorexia; trauma; asthma; eczema; nausea;
KW hypertension; neurological disorder; Parkinson's disease; drug screening;
KW cardiant; cell proliferative disorder; multiple sclerosis; osteoporosis;
KW diabetes mellitus; glomerulonephritis; cardiovascular disorder; anaemia;
KW autoimmune disorder; inflammatory disorder; myocardial infarction; AIDS;
KW developmental disorder; reproductive disorder; infertility; diarrhoea;
KW dementia; acidosis; cataract; gynaecomastia; epilepsy; jaundice; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   616..2358
FT   /*tag= a
FT   /product= "Human protease PRTS-2 protein"
XX
PN WO200183775-A2.
XX
PD 08-NOV-2001.
XX
PF 04-MAY-2001; 2001WO-US14651.
XX
PR 04-MAY-2000; 2000US-202082P.
PR 11-MAY-2000; 2000US-203566P.
PR 17-MAY-2000; 2000US-205803P.
PR 25-MAY-2000; 2000US-207477P.
PR 01-JUN-2000; 2000US-209402P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Delegeane AM, Lal P, Hafalia A, Patterson C, Walia NK, Kearney L;
PI Tribouley CM, Khan FA, Yao MG, Baughn MR, Azimzai Y, Elliott VS;
PI Nguyen DB, Gandhi AR, Yang J, Hernandez R, Policky JL, Lu DAM;
PI Reddy R, Yue H, Tang YT;
XX
DR WPI; 2002-034518/04.
DR P-PSDB; AAE14337.
XX
PT Novel human proteases and polynucleotides encoding the proteases,
PT useful for treating, diagnosing or preventing cell proliferative,
PT cardiovascular, autoimmune/inflammatory, neurological and developmental
PT disorders -
XX
PS Claim 5; Page 139-140; 151pp; English.
XX
CC The invention relates to human proteases (PRTS1-14) and its corresponding
CC cDNA molecules. Human PRTS and its nucleic acid molecule are useful for
CC the diagnosis, treatment and prevention of disorders associated with
CC increased or decreased expression of PRTS. Examples of such disorders
CC include, cell proliferative disorders (arteriosclerosis, atherosclerosis,
CC hepatitis, psoriasis and cancers); autoimmune/inflammatory disorders
CC (AIDS, Addison's disease, allergy, anaemia, asthma, atopic dermatitis,
CC diabetes mellitus, glomerulonephritis, multiple sclerosis, osteoporosis,
CC trauma, Grave's disease, rheumatoid arthritis, ulcerative colitis, and
CC viral, bacterial, fungal, parasitic, protozoal and helminthic
CC infections); cardiovascular disorders (myocardial infarction, ischaemic
CC heart disease and hypertension); neurological disorders (epilepsy,
CC Alzheimer's disease, Pick's disease, Huntington's disease, dementia,
CC Parkinson's disease, stroke, mental disorders including mood, anxiety
CC and seasonal affective disorder and prion diseases); gastrointestinal
CC disorders (Crohn's disease, anorexia, nausea, diarrhoea and jaundice);
CC epithelial disorders (contact dermatitis, eczema, acne vulgaris,
CC alopecia, scabies, insect bites and urticaria); reproductive disorder
CC (infertility, disruption of estrous and menstrual cycle and
CC gynaecomastia); and developmental disorders (renal tubular acidosis,
CC Cushing's syndrome, seizure disorders, congenital glaucoma and cataract).
CC PRTS DNA is also in useful is gene therapy. PRTS and its immunogenic
CC fragments are useful for screening libraries of compounds in several drug
CC screening assays. The present sequence is human protease PRTS-2 cDNA.
XX
SQ Sequence 2510 BP; 777 A; 494 C; 527 G; 712 T; 0 other;

Alignment Scores:

Pred. No.:	8.78e-168	Length:	2510
Score:	1522.00	Matches:	309
Percent Similarity:	58.08%	Conservative:	0
Best Local Similarity:	58.08%	Mismatches:	0
Query Match:	78.74%	Indels:	223
DB:	24	Gaps:	2

US-10-070-464-7 (1-360) x AAD23843 (1-2510)

```
Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        609 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 668

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        669 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 728

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        729 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 788

Qy         61 MetLeuGluThrArg-ArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPr 80
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        789 ATGTTGGAAACAAGGCAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCC 848

Qy         80 oLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle----- 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        849 TAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGT 908

Qy         98 ---------------------------------------------------------- 98

Db        909 CATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGC 968

Qy         98 ---------------------------------------------------------- 98

Db        969 CAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCA 1028

Qy         98 ---------------------------------------------------------- 98
```

```
Db       1029 GACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGT 1088

Qy         98 ------------------------------------------------------------ 98

Db       1089 TATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTA 1148

Qy         98 ------------------------------------------------------------ 98

Db       1149 TGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAG 1208

Qy         98 ------------------------------------------------------------ 98

Db       1209 TCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTT 1268

Qy         98 ------------------------------------------------------------ 98

Db       1269 ATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCC 1328

Qy         98 ------------------------------------------------------------ 98

Db       1329 TGCTCCAACTGTCACTTGGATGATCACATTCATGAGATCTCTAGGAACTCCATCCTGTAT 1388

Qy         99 ---------------------------ValAspGluValArgArgLeuValTyrPheGl 109
                                         |||||||||||||||||||||||||||||||||
Db       1389 GTGTGTGACACATATAGTTGAGATCCAAGTTGATGAAGTCAGAAGGCTGGTATATTTTGA 1448

Qy        109 uGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsnProGl 129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1449 AGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAATCCTGG 1508

Qy        129 yGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGlnHisCy 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1509 AGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAGCACTG 1568

Qy        149 sAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeuTyrLy 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1569 TGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTTTACAA 1628

Qy        169 sLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThrIleLe 189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1629 GCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACCATTTT 1688

Qy        189 uAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGluSerTh 209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1689 GGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAAAGTAC 1748

Qy        209 rThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGlyLysLy 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1749 TACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGAAAGAA 1808

Qy        229 sTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsnArgPh 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1809 ATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAATCGGTT 1868

Qy        249 eLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValValValVa 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1869 TAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTTGTAGT 1928

Qy        269 lIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLysTyrLy 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1929 GATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAATATAA 1988

Qy        289 sMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTy 309
              ||||
Db       1989 AATG-------------------------------------------------------- 1992

Qy        309 rAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSe 329

Db       1992 ------------------------------------------------------------ 1992

Qy        329 rLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProVa 349
                                                    ||||||||||||||||||||||
Db       1993 --------------------------------------GTTGCTATTGCTGGGGCCCAGT 2015

Qy        349 lThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||
Db       2016 CACTCTGTGGATCTTCTATGATACAGGATACACG 2049
```

```
RESULT 13
ABK83332
ID   ABK83332 standard; cDNA; 4685 BP.
XX
AC   ABK83332;
XX
DT   12-AUG-2002  (first entry)
XX
DE   cDNA encoding human DPRP-1 splice variant #8.
XX
KW   Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW   DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW   diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW   heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW   ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW   dyskinesia; reproductive disorder; inflammatory disorder;
KW   metabolic disorder; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200231134-A2.
XX
PD   18-APR-2002.
XX
PF   12-OCT-2001; 2001WO-US31874.
XX
PR   12-OCT-2000; 2000US-240117P.
XX
PA   (FERR ) FERRING BV.
XX
PI   Qi S,  Akinsanya KO,  Riviere PJ,  Junien J;
XX
DR   WPI; 2002-444178/47.
DR   P-PSDB; ABG61601.
XX
PT   New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT   the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT   viral infections, cancers, allergies, neurological disorders, or pain
PT   -
XX
PS   Disclosure; Page 75-76; 113pp; English.
XX
CC   The present invention relates to the isolation of novel human serine
CC   proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC   proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC   and nucleic acids encoding them are useful for treating infections
CC   such as fungal, bacterial, protozoan and viral infections, particularly
CC   infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC   pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC   bulimia, Parkinson's disease, acute heart failure, hypotension,
CC   hypertension, urinary retention, osteoporosis, angina pectoris,
CC   stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC   psychotic and neurological disorders (e.g. anxiety, dementia, or
CC   schizophrenia), and dyskinesias. These may also be used in discovering
CC   therapeutic agents for the treatment of reproductive, inflammatory and
CC   metabolic disorders. ABK83322-ABK83343 encode human DPRP proteins.
XX
SQ   Sequence 4685 BP; 1430 A; 853 C; 991 G; 1411 T; 0 other;

Alignment Scores:
Pred. No.:              2.26e-167     Length:          4685
Score:                  1522.00       Matches:         312
Percent Similarity:     58.54%        Conservative:    0
Best Local Similarity:  58.54%        Mismatches:      1
Query Match:            78.74%        Indels:          221
DB:                     24            Gaps:            2

US-10-070-464-7 (1-360) x ABK83332 (1-4685)
```

```
Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1002

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1062

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1122

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1182

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||------
Db       1183 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1242

Qy         98 ---------------------------------------------------------- 98

Db       1243 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1302

Qy         98 ---------------------------------------------------------- 98

Db       1303 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1362

Qy         98 ---------------------------------------------------------- 98

Db       1363 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1422

Qy         98 ---------------------------------------------------------- 98

Db       1423 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1482

Qy         98 ---------------------------------------------------------- 98

Db       1483 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1542

Qy         98 ---------------------------------------------------------- 98

Db       1543 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1602

Qy         98 ---------------------------------------------------------- 98

Db       1603 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1662

Qy         98 ---------------------------------------------------------- 98

Db       1663 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1722

Qy         99 ------------------------------ValAspGluValArgArgLeuValTyr 107
                                            |||||||||||||||||||||||||
Db       1723 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1782

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1842

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1902

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1962

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1963 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2022

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||
Db       2023 ATTTTGGATTCAGT---------------------------------------------- 2036

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227

Db       2036 ---------------------------------------------------------- 2036

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
                                                  ||||||||||||||||||||||||
Db       2037 ----------------------------------CCTCAGGTGCAGTTGGTGAATAAT 2060

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2061 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2120

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2121 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2180

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2181 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2240

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2241 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2300

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2301 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2360

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||||
Db       2361 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2399
```

RESULT 14
ABK83331
ID ABK83331 standard; cDNA; 4676 BP.
XX
AC ABK83331;
XX
DT 12-AUG-2002 (first entry)
XX
DE cDNA encoding human DPRP-1 splice variant #7.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI; 2002-444178/47.
DR P-PSDB; ABG61600.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and

```
PT   viral infections, cancers, allergies, neurological disorders, or pain
PT   -
XX
PS   Disclosure; Page 72-73; 113pp; English.
XX
CC   The present invention relates to the isolation of novel human serine
CC   proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC   proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC   and nucleic acids encoding them are useful for treating infections
CC   such as fungal, bacterial, protozoan and viral infections, particularly
CC   infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC   pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC   bulimia, Parkinson's disease, acute heart failure, hypotension,
CC   hypertension, urinary retention, osteoporosis, angina pectoris,
CC   stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC   psychotic and neurological disorders (e.g. anxiety, dementia, or
CC   schizophrenia), and dyskinesias. These may also be used in discovering
CC   therapeutic agents for the treatment of reproductive, inflammatory and
CC   metabolic disorders. ABK83322-ABK83343 encode human DPRP proteins.
XX
SQ   Sequence 4676 BP; 1424 A; 859 C; 979 G; 1414 T; 0 other;

Alignment Scores:
Pred. No.:                2.58e-167      Length:          4676
Score:                    1521.50        Matches:         309
Percent Similarity:       57.97%         Conservative:    0
Best Local Similarity:    57.97%         Mismatches:      1
Query Match:              78.71%         Indels:          224
DB:                       24             Gaps:            2

US-10-070-464-7 (1-360) x ABK83331 (1-4676)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1002

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1062

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1122

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1182

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1242

Qy         98 ---------------------------------------------------------- 98

Db       1243 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1302

Qy         98 ---------------------------------------------------------- 98

Db       1303 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1362

Qy         98 ---------------------------------------------------------- 98

Db       1363 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1422

Qy         98 ---------------------------------------------------------- 98

Db       1423 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1482

Qy         98 ---------------------------------------------------------- 98

Db       1483 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1542

Qy         98 ---------------------------------------------------------- 98

Db       1543 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1602

Qy         98 ---------------------------------------------------------- 98

Db       1603 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1662

Qy         98 ---------------------------------------------------------- 98

Db       1663 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1722

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                 ||||||||||||||||||||||||
Db       1723 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1782

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1842

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1902

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1962

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1963 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2022

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2023 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 2082

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2083 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 2142

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||
Db       2143 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCG---------------------- 2180

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267

Db       2180 ---------------------------------------------------------- 2180

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287

Db       2180 ---------------------------------------------------------- 2180

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2181 ---------GGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2231

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2232 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2291

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2292 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2351

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||||
Db       2352 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2390

RESULT 15
ABV76411
ID   ABV76411 standard; cDNA; 1821 BP.
XX
AC   ABV76411;
XX
```

```
DT   07-FEB-2003  (first entry)
XX
DE   Dipeptidyl peptidase IV protein 27.5-encoding cDNA.
XX
KW   Dipeptidyl peptidase IV protein 27.5; recombinant production;
KW   Stenotrophomonas maltophilia dipeptidyl peptidase IV homologue;
KW   gene therapy; diabetes; tumour; cancer; cytostatic; gene; ss.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             826..1578
FT                   /*tag=  a
FT                   /product= "Dipeptidyl peptidase IV protein 27.5"
XX
PN   CN1342768-A.
XX
PD   03-APR-2002.
XX
PF   12-SEP-2000; 2000CN-0125127.
XX
PR   12-SEP-2000; 2000CN-0125127.
XX
PA   (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX
PI   Mao Y,  Xie Y;
XX
DR   WPI; 2002-529809/57.
DR   P-PSDB; ABB99949.
XX
PT   New dipeptidyl peptidase IV protein 27.5 and encoding polynucleotide,
PT   useful for treating diabetes and tumors -
XX
PS   Claim 6; Page 24-25 (Disclosure); 32pp; Chinese.
XX
CC   The invention relates to dipeptidyl peptidase IV protein 27.5 (ABB99949)
CC   and nucleic acids encoding it (ABV76411). The protein has a molecular
CC   weight of 27.5 kD and has 40% identity and 58% homology over a 227 amino
CC   acid stretch with Stenotrophomonas maltophilia dipeptidyl peptidase IV
CC   (GenBank accession number D83263). The invention also relates to a method
CC   for the recombinant production of the protein, an antagonist of the
CC   protein, and the use of the protein, gene and antagonist in therapeutic
CC   applications. Dipeptidyl peptidase IV protein 27.5 can be used in the
CC   treatment of a variety of diseases such as diabetes and tumours. The
CC   present sequence represents cDNA encoding dipeptidyl peptidase IV
CC   protein 27.5.
XX
SQ   Sequence 1821 BP; 533 A; 374 C; 391 G; 523 T; 0 other;

Alignment Scores:
Pred. No.:              4.42e-157      Length:          1821
Score:                  1429.00        Matches:         279
Percent Similarity:     84.84%         Conservative:    12
Best Local Similarity:  81.34%         Mismatches:      22
Query Match:            73.93%         Indels:          30
DB:                     24             Gaps:            6

US-10-070-464-7 (1-360) x ABV76411 (1-1821)

Qy         43 LeuTyrGluGluAsnAspGluSerGluValGluIleIle---------------------- 55
              ::::::   ::::::   |||   |||:::|||   |||
Db        244 GTTTTTCCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACA 303

Qy         56 ------------------HisValThrSerProMetLeuGluThrArgArgAlaAspSer 69
                              :::||||||   :::   |||::::::
Db        304 GGTTTCCGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAA------------ 351

Qy         70 PheArgTyrProLysThrGlyThrAlaAsnProLysValThrPheLys---------Met 86
              ::::::             |||          |||          ||||||
Db        352 TATAAACGATCCAGTGGTGGGCTGCCTGCTCCAAGT---GATTTCAAGTGTCCTATCAAA 408

Qy         87 SerGluIleMetIle---------------AspAlaGluGlyArg------------Ile 97
                 ||||||   |||               :::      ||||||            |||
```

```
Db        409 GAGGAGATAGCAATTACCAGTGGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATC 468

Qy         98 IleValAspGluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGlu 117
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        469 CAAGTTGATGAAGTCAGAAGGCTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAG 528

Qy        118 HisHisLeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArg 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        529 CATCACCTGTACGTAGTCAGTTACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGT 588

Qy        138 GlyTyrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSer 157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        589 GGCTACTCACATTCTTGCTGCATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGT 648

Qy        158 AsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspPro 177
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        649 AACCAGAAGAATCCACACTGTGTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCA 708

Qy        178 ThrCysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAsp 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        709 ACTTGCAAAACAAAGGAATTTTGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGAC 768

Qy        198 TyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMet 217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        769 TATACTCCTCCAGAAATTTTCTCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATG 828

Qy        218 LeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyr 237
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        829 CTCTACAAGCCTCATGATCTACAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATAT 888

Qy        238 GlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeu 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        889 GGTGGTCCTCAGGTGCAGTTGGTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTG 948

Qy        258 AsnThrLeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCysHis 277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        949 AATACCCTAGCCTCTCTAGGTTATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCAC 1008

Qy        278 ArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAsp 297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1009 CGAGGGCTTAAATTTGAAGGCGCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGAT 1068

Qy        298 GlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgVal 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1069 CAGGTGGAAGGACTCCAATATCTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTG 1128

Qy        318 GlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSer 337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1129 GGCATCCACGGCTGGTCCTATGGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCA 1188

Qy        338 AspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThr 357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1189 GATATCTTCAGGGTTGCTATTGCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACA 1248

Qy        358 GlyTyrThr 360
              |||||||||
Db       1249 GGATACACG 1257
```

```
Search completed: October 15, 2003, 17:58:24
Job time : 292.512 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:29:02 ; Search time 56.4006 Seconds
(without alignments)
2817.309 Million cell updates/sec

Title: US-10-070-464-7
Perfect score: 1933
Sequence: 1 EEDARSAGVATFVLQEEFDR.....RVAIAGAPVTLWIFYDTGYT 360

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10070464/runat_15102003_113554_24845/app_query.fasta_1.2652
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10070464_@CGN_1_1_146_@runat_15102003_113554_24845 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	347.5	18.0	5496	4	US-09-462-284-1	Sequence 1, Appli
	2	346	17.9	823	4	US-09-280-116-171	Sequence 171, App
	3	315	16.3	2924	5	PCT-US93-07923-1	Sequence 1, Appli
	4	315	16.3	3407	4	US-10-002-593-5	Sequence 5, Appli
c	5	283	14.6	657	4	US-09-221-017B-646	Sequence 646, App
	6	257	13.3	2815	1	US-08-230-491A-1	Sequence 1, Appli
	7	257	13.3	2815	1	US-08-619-280A-1	Sequence 1, Appli
	8	257	13.3	2815	2	US-08-940-391-1	Sequence 1, Appli
	9	236.5	12.2	543	4	US-09-221-017B-253	Sequence 253, App
	10	233.5	12.1	535	4	US-09-389-681-428	Sequence 428, App
	11	233.5	12.1	535	4	US-09-620-405B-428	Sequence 428, App
	12	233.5	12.1	535	4	US-09-433-826B-428	Sequence 428, App
	13	233.5	12.1	535	4	US-09-604-287A-428	Sequence 428, App
c	14	211	10.9	4982	3	US-08-699-103B-1	Sequence 1, Appli
c	15	211	10.9	4982	4	US-09-229-059-1	Sequence 1, Appli
c	16	211	10.9	4982	4	US-09-628-133-1	Sequence 1, Appli
	17	204	10.6	1896	3	US-09-016-080-2	Sequence 2, Appli
c	18	190	9.8	612	4	US-09-392-184-31	Sequence 31, Appl
c	19	185	9.6	3085	4	US-09-221-017B-1010	Sequence 1010, Ap
	20	156.5	8.1	1869	2	US-08-664-646A-1	Sequence 1, Appli
	21	156.5	8.1	1869	2	US-09-066-285-1	Sequence 1, Appli
	22	156.5	8.1	1869	3	US-09-261-006-1	Sequence 1, Appli
	23	156.5	8.1	1869	3	US-08-951-088-1	Sequence 1, Appli
	24	156.5	8.1	1869	4	US-09-609-566-1	Sequence 1, Appli
	25	156.5	8.1	1869	4	US-09-609-570-1	Sequence 1, Appli
	26	156.5	8.1	1869	4	US-09-427-372-1	Sequence 1, Appli
	27	156.5	8.1	1869	4	US-09-693-554-1	Sequence 1, Appli
	28	136	7.0	904	4	US-09-221-017B-58	Sequence 58, Appl
	29	132.5	6.8	1032	4	US-09-724-623-17	Sequence 17, Appl
	30	129	6.7	815	4	US-09-221-017B-962	Sequence 962, App
c	31	116.5	6.0	675	4	US-09-252-991A-3444	Sequence 3444, Ap
	32	116.5	6.0	1536	4	US-09-252-991A-3483	Sequence 3483, Ap
	33	116.5	6.0	1845	4	US-09-252-991A-3489	Sequence 3489, Ap
	34	108.5	5.6	1470	1	US-08-589-893-11	Sequence 11, Appl
	35	108.5	5.6	1470	2	US-09-020-991-11	Sequence 11, Appl
	36	108.5	5.6	1470	2	US-09-062-890-11	Sequence 11, Appl
	37	106.5	5.5	1470	2	US-09-062-890-37	Sequence 37, Appl
	38	105	5.4	1365	4	US-09-252-991A-8265	Sequence 8265, Ap
	39	105	5.4	1590	4	US-09-252-991A-8140	Sequence 8140, Ap
	40	104.5	5.4	1710	3	US-09-068-960-14	Sequence 14, Appl
c	41	104	5.4	1230025	4	US-09-198-452A-1	Sequence 1, Appli
c	42	104	5.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	43	104	5.4	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	44	102.5	5.3	1470	1	US-08-124-674-1	Sequence 1, Appli
	45	102.5	5.3	1470	1	US-08-589-893-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-462-284-1
; Sequence 1, Application US/09462284
; Patent No. 6309868
; GENERAL INFORMATION:
; APPLICANT: Nestec S.A.
; APPLICANT: Monod, Michel
; APPLICANT: Doumas, Agnes
; APPLICANT: Affolter, Micheal
; APPLICANT: Van Den Broek, Peter
; TITLE OF INVENTION: CLONING OF THE
; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM
; TITLE OF INVENTION: ASPERGILLUS ORYZAE
; FILE REFERENCE: 8265-298
; CURRENT APPLICATION NUMBER: US/09/462,284
; CURRENT FILING DATE: 2000-01-03
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5496
; TYPE: DNA
; ORGANISM: Fungus
US-09-462-284-1

Alignment Scores:

Pred. No.:	1.45e-30	Length:	5496
Score:	347.50	Matches:	124
Percent Similarity:	37.09%	Conservative:	57
Best Local Similarity:	25.41%	Mismatches:	150
Query Match:	17.98%	Indels:	157
DB:	4	Gaps:	16

US-10-070-464-7 (1-360) x US-09-462-284-1 (1-5496)

Qy 8 GlyValAlaThrPheValLeuGlnGluGluPhe------AspArgTyrSerGlyTyrTrp 25

```
Db       2486 GGCGTGCCGGACTGGATCTATGAAGAGGAGATCCTCGGCGATCGCTACGCG---TTGTGG 2542

Qy         26 TrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeuArgIleLeuTyrGlu 45

Db       2543 TTCTCGCCAGATGGTGAA-------------------------TATCTGGCTTACTTG 2575

Qy         46 GluAsnAspGluSerGluValGluIleIleHisValThrSerProMetLeuGluThrArg 65

Db       2576 AGCTTCAATGAGACTGGGGTTCCGACCTACACCGTTCAGTATTATATGGATAACCAAGAG 2635

Qy         66 ArgAlaAspSerPhe-----------------ArgTyrProLysThrGlyThrAlaAsn 79

Db       2636 ATCGCTCCGGCGTATCCATGGGAGCTGAAGATAAGGTATCCCAAGGTGTCGCAGACGAAT 2695

Qy         80 ProLysValThr-------------------------------------------------- 83

Db       2696 CCGACCGTGACGTTGAGTCTGCTTAACATCGCTAGCAAGGAGGTGAAGCAGGCGCCGATC 2755

Qy         84 ------PheLysMetSerGluIleMetIle---------------------------- 91

Db       2756 GACGCGTTCGAGTCAACTGACTTGATCATTGGCGAGGTTGCTTGGCTCACTGATACTCAC 2815

Qy         91 -------------------------------------------------------------- 91

Db       2816 ACCACCGTTGCTGCTAAGGCGTTCAACCGTGTCCAGGACCAGCAAAAGGTCGTCGCGGTC 2875

Qy         92 ---------------------------------AspAlaGluGlyArg------------ 96

Db       2876 GATACTGCCTCGAACAAGGCTACTGTCATCAGCGACCGAGATGGGACCGATGGATGGCTC 2935

Qy         96 -------------------------------------------------------------- 96

Db       2936 GATAACCTTCTTTCAATGAAGTATATTGGCCCTATCAAGCCGTCCGACAAGGATGCCTAC 2995

Qy         96 -------------------------------------------------------------- 96

Db       2996 TACATCGACATCTCTGACCATTCGGGATGGGCGCATCTGTATCTCTTCCCCGTTTCGGGC 3055

Qy         97 ---------------------------------------------------IleIleValAsp 100

Db       3056 GGCGAACCTATCCCACTAACCAAAGGCGACTGGGAGGTCACGTCTATTCTGAGTATTGAT 3115

Qy        101 GluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeu 120

Db       3116 CAGGAACGCCAGTTGGTGTACTACCTGTCGACTCAACACCACAGCACCGAGCGCCATCTC 3175

Qy        121 TyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSer 140

Db       3176 TACTCCGTCTCCTAT---TCCACGTTTGCGGTCACCCCGCTCGTCGACGACACCGTTGCC 3232

Qy        141 ------HisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsn 158

Db       3233 GCGTACTGGTCTGCTTCCTTCTCCGCGAACTCGGGCTACTACATCCTCACATACGGAGGC 3292

Qy        159 GlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThr 178

Db       3293 CCAGACGTACCCTAC---CAGGAACTCTACACGACCAACAGT----------------- 3331

Qy        179 CysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeu--------- 195

Db       3332 ------ACCAAACCACTC---CGCACAATCACCGACAACGCCAAAGTACTCGAGCAAATC 3382

Qy        196 ProAspTyrThrProProGluIlePheSerPheGlu------SerThrThrGlyPheThr 213

Db       3383 AAGGACTATGCATTGCCCAACATCACCTACTTCGAGCTTCCCCTCCCCTCCGGAGAAACC 3442

Qy        214 LeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrVal 233

Db       3443 CTCAATGTGATGCAGCGCTTACCCCCCGGGTTCTCCCCGGATAAGAAGTACCCCATACTT 3502

Qy        234 LeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLys 253

Db       3503 TTCACCCCATACGGCGGCCCAGGCGCCCAAGAAGTGACCAAGAGATGGCAAGCCCTGAAT 3562
```

```
Qy        254 Tyr---PheArgLeuAsnThrLeuAlaSerLeuGlyTyrValValValValIleAspAsn 272

Db       3563 TTCAAGGCCTATGTCGCCTCCGACAGCGAACTCGAGTACGTAACCTGGACTGTCGACAAC 3622

Qy        273 ArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGln 292

Db       3623 CGCGGCACAGGTTTCAAAGGACGCAAGTTCCGCTCCGCCGTCACGCGCCAACTCGGCCTC 3682

Qy        293 IleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIle 312

Db       3683 CTCGAAGCAGAAGACCAGATCTACGCCGCGCAACAG---GCGGCCAACATCCCCTGGATC 3739

Qy        313 AspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAla 332

Db       3740 GATGCAGACCACATCGGCATCTGGGGCTGGAGTTTCGGAGGCTACTTGACCAGCAAGGTC 3799

Qy        333 LeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrp 352

Db       3800 CTGGAGAAGGACAGCGGTGCTTTCACATTAGGAGTCATCACCGCCCCTGTTTCTGACTGG 3859

Qy        353 IlePheTyrAspThrGlyTyrThr 360

Db       3860 CGTTTCTACGACTCAATGTACACG 3883
```

RESULT 2
US-09-280-116-171
; Sequence 171, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 171
; LENGTH: 823
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: prolyl oligopeptidases
US-09-280-116-171

Alignment Scores:

Pred. No.:	1.09e-31	Length:	823
Score:	346.00	Matches:	67
Percent Similarity:	98.57%	Conservative:	2
Best Local Similarity:	95.71%	Mismatches:	0
Query Match:	17.90%	Indels:	1
DB:	4	Gaps:	0

US-10-070-464-7 (1-360) x US-09-280-116-171 (1-823)

```
Qy        292 GlnIleGluIleAspAspGlnValGluGly-LeuGlnTyrLeuAlaSerArgTyrAspPh 311

Db          3 CAAATAGAAATTGACGATCAGGTGGAAGGACATCCAATATCTAGCTTCTCGATATGATTT 62

Qy        311 eIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMe 331

Db         63 CATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATACCTCTCCCTGAT 122

Qy        331 tAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaProValThrLe 351

Db        123 GGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGTCCCCAGTCACTCT 182

Qy        351 uTrpIlePheTyrAspThrGlyTyrThr 360

Db        183 GTGGATCTTCTATGATACAGGATACACG 210
```

RESULT 3
PCT-US93-07923-1

```
; Sequence 1, Application PC/TUS9307923
;  GENERAL INFORMATION:
;    APPLICANT:  Morimoto, Chikao
;    APPLICANT:  Schlossman, Stuart F.
;    APPLICANT:  Tanaka, Toshiaki
;    TITLE OF INVENTION:  HUMAN CD26 AND METHODS FOR USE
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;      SOFTWARE:  WordPerfect (Version 5.0)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/07923
;      FILING DATE:  19930819
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/934,162
;      FILING DATE:  21-AUG-1992
;      APPLICATION NUMBER:  07/832,211
;      FILING DATE:  06-FEB-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fraser, Janis K.
;      REGISTRATION NUMBER:  34,819
;      REFERENCE/DOCKET NUMBER:  00530/055002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2924
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
PCT-US93-07923-1

Alignment Scores:
Pred. No.:              3.99e-27      Length:        2924
Score:                  315.00        Matches:       120
Percent Similarity:     34.89%        Conservative:  59
Best Local Similarity:  23.39%        Mismatches:    146
Query Match:            16.30%        Indels:        188
DB:                     5             Gaps:          22

US-10-070-464-7 (1-360) x PCT-US93-07923-1 (1-2924)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu---PheAsp 19
Db        578 AAAGAAGATATAATATATAATGGAATAACTGACTGGGTTTATGAAGAGGAAGTCTTCAGT 637

Qy         20 ArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIle 39
Db        638 GCCTACTCTGCTCTGTGGTGGTCTCCAAACGGCACTTTT--------------------- 676

Qy         40 LeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHis--------- 56
Db        677 ------TTAGCATATGCCCAATTTAACGACACAGAAGTCCCACTTATTGAATACTCCTTC 730

Qy         57 ValThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGly 76
Db        731 TACTCTGATGAGTCACTGCAGTACCCAAAGACTGTACGGGTTCCATATCCAAAGGCAGGA 790

Qy         77 ThrAlaAsnProLysValThrPhe------------------------------------ 84
Db        791 GCTGTGAATCCAACTGTAAAGTTCTTTGTTGTAAATACAGACTCTCTCAGCTCAGTCACC 850

Qy         85 LysMetSerGluIleMetIleAspAlaGluGlyArgIleIleValAspGluValArgArg 104
Db        851 AATGCAACTTCCATACAAATCACTGCTCCTGCTTCTATGTTGATAGGG------------ 898

Qy        105 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 124
Db        899 ------------------------------------GATCACTACTTGTGTGATGTGACA 922

Qy        125 TyrValAsnProGlyGluValThr---------------ArgLeuThrAspArgGlyTyr 139
Db        923 TGGGCAACACAAGAAAGAATTTCTTTGCAGTGGCTCAGGAGGATTCAGAACTATTCGGTC 982

Qy        140 SerHisSerCys-----------------------------CysIle------SerGln 147
Db        983 ATGGATATTTGTGACTATGATGAATCCAGTGGAAGATGGAACTGCTTAGTGGCACGGCAA 1042

Qy        148 HisCysAspPhe-----------PheIleSerLysTyrSerAsnGlnLysAsnProHis 163
Db       1043 CACATTGAAATGAGTACTACTGGCTGGGTTGGAAGATTT---AGGCCTTCAGAACCTCAT 1099

Qy        164 CysVal-----------SerLeuTyrLysLeuSerSerProGluAsp------------ 175
Db       1100 TTTACCCTTGATGGTAATAGCTTCTACAAGATCATCAGCAATGAAGAAGGTTACAGACAC 1159

Qy        176 --------------------------AspProThrCysLysThrLysGluPheTrpAla 186
Db       1160 ATTTGCTATTTCCAAATAGATAAAAAAGACTGCACATTTATTACAAAAGGCACCTGGGAA 1219

Qy        187 ThrIle------------------------------------------------------ 188
Db       1220 GTCATCGGGATAGAAGCTCTAACCAGTGATTATCTATACTACATTAGTAATGAATATAAA 1279

Qy        188 ------------------------------------------------------------ 188
Db       1280 GGAATGCCAGGAGGAAGGAATCTTTATAAAATCCAACTTAGTGACTATACAAAAGTGACA 1339

Qy        188 ------------------------------------------------------------ 188
Db       1340 TGCCTCAGTTGTGAGCTGAATCCGGAAAGGTGTCAGTACTATTCTGTGTCATTCAGTAAA 1399

Qy        189 -----------------LeuAspSerAlaGlyPro---LeuProAspTyrThr------ 199
Db       1400 GAGGGGAAGTATTATCAGCTGAGATGTTCCGGTCCTGGTCTGCCCCTCTATACTCTACAC 1459

Qy        199 ------------------------------------------------------------ 199
Db       1460 AGCAGCGTGAATGATAAAGGGCTGAGAGTCCTGGAAGACAATTCAGCTTTGGATAAAATG 1519

Qy        200 -----------------ProProGluIlePheSerPhe------GluSerThrThrGly 211
Db       1520 CTGCAGAATGTCCAGATGCCCTCCAAAAAACTGGACTTCATTATTTTGAATGAAACAAAA 1579

Qy        212 PheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrPro 231
Db       1580 TTT---TGGTATCAGATGATCTTGCCTCCTCAT---TTTGATAAATCCAAGAAATATCCT 1633

Qy        232 ThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGly 251
Db       1634 CTACTATTAGATGTGTATGCAGGCCCATGTAGTCAA-----------------AAAGCA 1675

Qy        252 ValLysTyrPheArgLeuAsn---------ThrLeuAlaSerLeuGlyTyrValValVal 268
Db       1676 GACACTGTCTTCAGACTGAACTGGGCCACTTACCTTGCAAGCACAGAAAACATTATAGTA 1735

Qy        269 Val---IleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
Db       1736 GCTAGCTTTGATGGCAGAGGAAGTGGTTACCAAGGAGATAAGATCATGCATGCAATCAAC 1795

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
Db       1796 AGAAGACTGGGAACATTTGAAGTTGAAGATCAAATTGAAGCAGCCAGACAATTT---TCA 1852
```

```
Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
Db       1853 AAAATGGGATTTGTGGACAACAAACGAATTGCAATTTGGGGCTGGTCATATGGAGGGTAC 1912

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
Db       1913 GTAACCTCAATGGTCCTGGGATCAGGAAGTGGCGTGTTCAAGTGTGGAATAGCCGTGGCG 1972

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
Db       1973 CCTGTATCCCGGTGGGAGTACTATGACTCAGTGTACACA 2011

RESULT 4
US-10-002-593-5
; Sequence 5, Application US/10002593
; Patent No. 6586198
; GENERAL INFORMATION:
;  APPLICANT: Vanderbilt University
;  APPLICANT:  Brown, Nancy J.
;  TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOTENSIN CONVERTIN
;  TITLE OF INVENTION:  INHIBITOR AND VASOPEPTIDASE INHIBITOR ASSOCIATED ANGIOEDEMA
;  FILE REFERENCE: Atty Docket No. 6586198 1242/48/2
;  CURRENT APPLICATION NUMBER: US/10/002,593
;  CURRENT FILING DATE:  2001-10-31
;  PRIOR APPLICATION NUMBER: 60/244,524
;  PRIOR FILING DATE: 2000-10-31
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;   LENGTH: 3407
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-002-593-5

Alignment Scores:
Pred. No.:                5.08e-27        Length:         3407
Score:                    315.00          Matches:        120
Percent Similarity:       34.89%          Conservative:   59
Best Local Similarity:    23.39%          Mismatches:     146
Query Match:              16.30%          Indels:         188
DB:                       4               Gaps:           22

US-10-070-464-7 (1-360) x US-10-002-593-5 (1-3407)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu---PheAsp 19
Db        643 AAAGAAGATATAATATATAATGGAATAACTGACTGGGTTTATGAAGAGGAAGTCTTCAGT 702

Qy         20 ArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIle 39
Db        703 GCCTACTCTGCTCTGTGGTGGTCTCCAAACGGCACTTTT---------- 741

Qy         40 LeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHis--------- 56
Db        742 ------TTAGCATATGCCCAATTTAACGACACAGAAGTCCCACTTATTGAATACTCCTTC 795

Qy         57 ValThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGly 76
Db        796 TACTCTGATGAGTCACTGCAGTACCCAAAGACTGTACGGGTTCCATATCCAAAGGCAGGA 855

Qy         77 ThrAlaAsnProLysValThrPhe---------------------------------- 84
Db        856 GCTGTGAATCCAACTGTAAAGTTCTTTGTTGTAAATACAGACTCTCTCAGCTCAGTCACC 915

Qy         85 LysMetSerGluIleMetIleAspAlaGluGlyArgIleIleValAspGluValArgArg 104
Db        916 AATGCAACTTCCATACAAATCACTGCTCCTGCTTCTATGTTGATAGGG------------ 963

Qy        105 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 124
Db        964 ----------------------------------GATCACTACTTGTGTGATGTGACA 987

Qy        125 TyrValAsnProGlyGluValThr--------------ArgLeuThrAspArgGlyTyr 139
Db        988 TGGGCAACACAAGAAAGAATTTCTTTGCAGTGGCTCAGGAGGATTCAGAACTATTCGGTC 1047

Qy        140 SerHisSerCys----------------------------CysIle------SerGln 147
Db       1048 ATGGATATTTGTGACTATGATGAATCCAGTGGAAGATGGAACTGCTTAGTGGCACGGCAA 1107

Qy        148 HisCysAspPhe------------PheIleSerLysTyrSerAsnGlnLysAsnProHis 163
Db       1108 CACATTGAAATGAGTACTACTGGCTGGGTTGGAAGATTT---AGGCCTTCAGAACCTCAT 1164

Qy        164 CysVal------------SerLeuTyrLysLeuSerSerProGluAsp------------ 175
Db       1165 TTTACCCTTGATGGTAATAGCTTCTACAAGATCATCAGCAATGAAGAAGGTTACAGACAC 1224

Qy        176 --------------------------AspProThrCysLysThrLysGluPheTrpAla 186
Db       1225 ATTTGCTATTTCCAAATAGATAAAAAAGACTGCACATTTATTACAAAAGGCACCTGGGAA 1284

Qy        187 ThrIle---------------------------------------------------- 188
Db       1285 GTCATCGGGATAGAAGCTCTAACCAGTGATTATCTATACTACATTAGTAATGAATATAAA 1344

Qy        188 ---------------------------------------------------------- 188
Db       1345 GGAATGCCAGGAGGAAGGAATCTTTATAAAATCCAACTTATTGACTATACAAAAGTGACA 1404

Qy        188 ---------------------------------------------------------- 188
Db       1405 TGCCTCAGTTGTGAGCTGAATCCGGAAAGGTGTCAGTACTATTCTGTGTCATTCAGTAAA 1464

Qy        189 ------------------LeuAspSerAlaGlyPro---LeuProAspTyrThr------ 199
Db       1465 GAGGCGAAGTATTATCAGCTGAGATGTTCCGGTCCTGGTCTGCCCCTCTATACTCTACAC 1524

Qy        199 ---------------------------------------------------------- 199
Db       1525 AGCAGCGTGAATGATAAAGGGCTGAGAGTCCTGGAAGACAATTCAGCTTTGGATAAAATG 1584

Qy        200 -----------------ProProGluIlePheSerPhe------GluSerThrThrGly 211
Db       1585 CTGCAGAATGTCCAGATGCCCTCCAAAAAACTGGACTTCATTATTTTGAATGAAACAAAA 1644

Qy        212 PheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrPro 231
Db       1645 TTT---TGGTATCAGATGATCTTGCCTCCTCAT---TTTGATAAATCCAAGAAATATCCT 1698

Qy        232 ThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGly 251
Db       1699 CTACTATTAGATGTGTATGCAGGCCCATGTAGTCAA-----------------AAAGCA 1740

Qy        252 ValLysTyrPheArgLeuAsn---------ThrLeuAlaSerLeuGlyTyrValValVal 268
Db       1741 GACACTGTCTTCAGACTGAACTGGGCCACTTACCTTGCAAGCACAGAAAACATTATAGTA 1800

Qy        269 Val---IleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
Db       1801 GCTAGCTTTGATGGCAGAGGAAGTGGTTACCAAGGAGATAAGATCATGCATGCAATCAAC 1860

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
Db       1861 AGAAGACTGGGAACATTTGAAGTTGAAGATCAAATTGAAGCAGCCAGACAATTT---TCA 1917

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
Db       1918 AAAATGGGATTTGTGGACAACAAACGAATTGCAATTTGGGGCTGGTCATATGGAGGGTAC 1977

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
Db       1978 GTAACCTCAATGGTCCTGGGATCGGGAAGTGGCGTGTTCAAGTGTGGAATAGCCGTGGCG 2037

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
Db       2038 CCTGTATCCCGGTGGGAGTACTATGACTCAGTGTACACA 2076
```

```
Db       2038 CCTGTATCCCGGTGGGAGTACTATGACTCAGTGTACACA 2076

RESULT 5
US-09-221-017B-646/c
; Sequence 646, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  646:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  657 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1...657
US-09-221-017B-646

Alignment Scores:
Pred. No.:              2.45e-24      Length:          657
Score:                  283.00        Matches:         60
Percent Similarity:     55.21%        Conservative:    30
Best Local Similarity:  36.81%        Mismatches:      69
Query Match:            14.64%        Indels:          4
DB:                     4             Gaps:            2

US-10-070-464-7 (1-360) x US-09-221-017B-646 (1-657)

Qy        198 TyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMet 217
              |||   |||   |||       :::   ::::::   :::|||       |||
Db        566 TACAACCCGAAGGAGTTTACCACTATCAAAACTCAATCGGGTCTTGAACTGAATGCCTGG 507

Qy        218 LeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyr 237
              :::   ||||||   |||       |||   :::   ||||||   ::::::      |||
Db        506 ATCGTGAAGCCTATTGATTTCGATCCCTCTCGCCACTATCCTGTCCTGATGGTACAGTAT 447

Qy        238 GlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeu 257
                ||||||       |||   |||   :::|||:::           :::
Db        446 AGCGGTCCCAACTCCCAGCAGGTATTGGATCGCTAT---------TCATTCGATTGGGAA 396

Qy        258 AsnThrLeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCysHis 277
              :::   |||||||||   ||||||||||||       :::|||   ||||||:::
Db        395 CACTACCTTGCATCGAAAGGTTACGTCGTGGCATGTGTGGATGGGCGTGGCACCGGTGCT 336

Qy        278 ArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAsp 297
              ||||||   ::::::                ::::::|||       |||   ||||||
Db        335 CGCGGCGAAGAATGGCGCAAGTGTACCTACATGCAACTCGGTGTATTCGAAAGCGATGAT 276

Qy        298 GlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgVal 317
              |||:::           :::       :::       ::::::|||       |||:::
Db        275 CAGATAGCAGCGGCCACTGCTATA---GGACAGCTGCCCTATGTGGATGCAGCTCGTATC 219

Qy        318 GlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSer 337
              ||||||   ||||||||||||||||||||   :::|||||||:::|||   :::   :::
Db        218 GGCATATGGGGGTGGAGCTATGGCGGCTATACCACACTAATGAGTTTGTGTCGGGGAAAT 159

Qy        338 AspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThr 357
              |||:::       ||||||   |||||||||       |||   |||||||||:::
Db        158 GGTACATTCAAAGCGGGGATAGCCGTTGCTCCTGTGGCAGACTGGCGTTTCTACGATTCG 99

Qy        358 GlyTyrThr 360
              ||||||
Db         98 GTTTACACC 90

RESULT 6
US-08-230-491A-1
; Sequence 1, Application US/08230491A
; Patent No. 5587299
;  GENERAL INFORMATION:
;    APPLICANT:  Rettig, Wolfgang J.; Scanlan, Matthew J.;
;    APPLICANT:  Garin-Chesa, Pilar; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULE CODING FOR
;    TITLE OF INVENTION:  FIBROBLAST ACTIVATION PROTEIN ` AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FELFE & LYNCH
;      STREET:  805 THIRD AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE 3.5 inch 1.2 MB STORAGE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  WORDPERFECT - ASC II
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/230,491A
;      FILING DATE:  20-APRIL-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5587299man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 330
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2815 Base pairs
;      TYPE:  nucleic acid
```

```
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-230-491A-1

Alignment Scores:
Pred. No.:              3.05e-20        Length:         2815
Score:                  257.00          Matches:        114
Percent Similarity:     36.70%          Conservative:   64
Best Local Similarity:  23.51%          Mismatches:     158
Query Match:            13.30%          Indels:         150
DB:                     1               Gaps:           20

US-10-070-464-7 (1-360) x US-08-230-491A-1 (1-2815)

Qy          8 GlyValAlaThrPheValLeuGlnGluGluPhe------AspArgTyrSerGlyTyrTrp 25
              |||:::      :::|||   :::||||||          :::|||:::      |||
Db        791 GGAATCCCAGACTGGGTTTATGAAGAGGAAATGCTTCCTACAAAATATGCT---CTCTGG 847

Qy         26 TrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeuArgIleLeuTyrGlu 45
              |||   |||                        ||||||   |||         |||
Db        848 TGGTCTCCT--------------------AATGGAAAATTTTTG------GCATATGCG 880

Qy         46 GluAsnAspGluSerGluValGluIleIleHisValThrSerProMetLeuGluThr--- 64
              |||   ::::::   ::::::   :::|||      :::             |||
Db        881 GAATTTAATGATAAGGATATACCAGTTATTGCCTATTCCTATTATGGCGATGAACAATAT 940

Qy         65 ArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnProLysValThrPhe 84
               |||      :::      ||||||||   |||      ||||||   |||
Db        941 CCTAGAACAATAAATATTCCATACCCAAAGGCTGGAGCTAAGAATCCCGTTGTTCGGATA 1000

Qy         85 LysMet---------------------------------SerGluIleMetIleAspAla 93
              :::                                    |||:::   :::   |||
Db       1001 TTTATTATCGATACCACTTACCCTGCGTATGTAGGTCCCCAGGAAGTGCCTGTTCCAGCA 1060

Qy         94 ------------------------------------------------GluGlyArgIle 97
                                                              :::   |||:::
Db       1061 ATGATAGCCTCAAGTGATTATTATTTCAGTTGGCTCACGTGGGTTACTGATGAACGAGTA 1120

Qy         98 IleValAspGluValArgArgLeuValTyrPheGlu------------------------ 109
              :::      ::::::|||:::
Db       1121 TGTTTGCAGTGGCTAAAAAGAGTCCAGAATGTTTCGGTCCTGTCTATATGTGACTTCAGG 1180

Qy        110 ---------GlyThrLysAspSerPro---LeuGluHisHisLeuTyr------------ 121
                       |||   |||   |||      :::   |||:::
Db       1181 GAAGACTGGCAGACATGGGATTGTCCAAAGACCCAGGAGCATATAGAAGAAAGCAGAACT 1240

Qy        122 --------------------------------ValValSerTyrValAsnProGlyGlu 130
                                             |||   ||||||
Db       1241 GGATGGGCTGGTGGATTCTTTGTTTCAAGACCAGTTTTCAGCTATGATGCCATTTCGTAC 1300

Qy        131 ValThrArgLeuThrAspArg---GlyTyrSerHisSerCysCysIleSerGlnHisCys 149
                  :::|||:::   ||||||   |||                :::||||||
Db       1301 TACAAAATATTTAGTGACAAGGATGGCTACAAACATAT-TCACTATATCAAAGACACTGT 1359

Qy        150 AspPhePheIleSerLysTyrSerAsnGln----------------------------- 159
                     |||   |||      |||
Db       1360 GGAAAATGCTATTCAAATTACAAGTGGCAAGTGGGAGGCCATAAATATATTCAGAGTAAC 1419

Qy        160 ------------------------------------------------LysAsnProHis 163
                                                             |||      |||
Db       1420 ACAGGATTCACTGTTTTATTCTAGCAATGAATTTGAAGAATACCCTGGAAGAAGAAACAT 1479

Qy        164 CysValSer---LeuTyrLysLeuSerSerProGluAsp-AspProThrCysLysThrLy 182
                  :::      :::||||||||||||   ::::::       ||||||      ::
Db       1480 CTACAGAATTAGCATTGGAAGCTATCCTCCAAGCAAGAAGTGTGTTACTTGCCATCTAAG 1539

Qy        182 sGlu--------------PheTrpAlaThrIleLeuAspSerAla-------------- 192
              ::::              :::   |||:::      |||   |||
Db       1540 GAAAGAAAGGTGCCAATATTACACAGCAAGTTTCAGCGACTACGCCAAGTACTATGCACT 1599

Qy        193 ----------GlyPro---LeuProAspTyrThr------------------------- 199
                      ||||||   :::|||      |||
Db       1600 TGTCTGCTACGGCCCAGGCATCCCCATTTCCACCCTTCATGATGGACGCACTGATCAAGA 1659

Qy        200 ------------------------------------------------------------Pr 200
                                                                           ||
Db       1660 AATTAAAATCCTGGAAGAAAACAAGGAATTGGAAAATGCTTTGAAAAATATCCAGCTGCC 1719

Qy        200 oProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLy 220
               |   |||                          ||||||:::      :::
Db       1720 TAAAGAGGAAATTAAGAAACTTGAAGTAGATGAAATTACTTTATGGTACAAGATGATTCT 1779

Qy        220 sProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPr 240
               |||                  ||||||||||||   ::::::   :::||||||||||||
Db       1780 TCCTCCTCAATTTGACAGATCAAAGAAGTATCCCTTGCTAATTCAAGTGTATGGTGGTCC 1839

Qy        240 oGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLe 260
               |      |||   |||   :::                  |||   :::|||   ::
Db       1840 CTGCAGTCAGAGTGTAAGGTCT-----------------GTATTTGCTGTTAATTGGAT 1881

Qy        260 uAlaSerLeu------------GlyTyrValValValValIleAspAsnArgGlySerCy 276
               ::::   |||           |||   |||:::   ::::::|||   ||||||:::
Db       1882 ATCTTATCTTGCAAGTAAGGAAGGGATGGTCATTGCCTTGGTGGATGGTCGAGGAACAGC 1941

Qy        276 sHisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAs 296
                  :::|||   |||          |||          |||:::|||      |||:::::
Db       1942 TTTCCAAGGTGACAAACTCCTCTATGCAGTGTATCGAAAGCTGGGTGTTTATGAAGTTGA 2001

Qy        296 pAspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspAr 316
               :||||||:::      ::::::                  |||||||||      ||
Db       2002 AGACCAGATTACAGCTGTCAGAAAATTCATA---GAAATGGGTTTCATTGATGAAAAAAG 2058

Qy        316 gValGlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnAr 336
               |:::   |||   ||||||||||||||||||||:::|||   :::||||||
Db       2059 AATAGCCATATGGGGCTGGTCCTATGGAGGATACGTTTCATCACTGGCCCTTGCATCTGG 2118

Qy        336 gSerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAs 356
                  :::   :::|||:::      ||||||   |||||||||:::   |||   ::|||
Db       2119 AACTGGTCTTTTCAAATGTGGTATAGCAGTGGCTCCAGTCTCCAGCTGGGAATATTACGC 2178

Qy        356 pThrGlyTyrThr 360
                  :::   ||||||
Db       2179 GTCTGTCTACACA 2191

RESULT 7
US-08-619-280A-1
; Sequence 1, Application US/08619280A
; Patent No. 5767242
;  GENERAL INFORMATION:
;    APPLICANT:  Zimmermann, Rainer; Park, John E.;
;    APPLICANT:  Rettig, Wolfgang; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
;    TITLE OF INVENTION:  ALPHA, AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 2.0 MB  storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/619,280A
;      FILING DATE:  18-MARCH-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/230,491
;      FILING DATE:  20-APRIL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5767242man D.
;      REGISTRATION NUMBER:  30,946
```

```
;      REFERENCE/DOCKET NUMBER:  LUD 5330.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2815 Base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-619-280A-1

Alignment Scores:
Pred. No.:                3.05e-20        Length:          2815
Score:                    257.00          Matches:         114
Percent Similarity:       36.70%          Conservative:    64
Best Local Similarity:    23.51%          Mismatches:      158
Query Match:              13.30%          Indels:          150
DB:                       1               Gaps:            20

US-10-070-464-7 (1-360) x US-08-619-280A-1 (1-2815)

Qy          8 GlyValAlaThrPheValLeuGlnGluGluPhe------AspArgTyrSerGlyTyrTrp 25
              |||:::    :::|||   :::||||||         :::|||:::     |||
Db        791 GGAATCCCAGACTGGGTTTATGAAGAGGAAATGCTTCCTACAAAATATGCT---CTCTGG 847

Qy         26 TrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeuArgIleLeuTyrGlu 45
              |||   |||                        ||||||   |||        |||
Db        848 TGGTCTCCT--------------------AATGGAAAATTTTTG------GCATATGCG 880

Qy         46 GluAsnAspGluSerGluValGluIleIleHisValThrSerProMetLeuGluThr--- 64
              |||   ::::::   ::::::   :::|||      :::              |||
Db        881 GAATTTAATGATAAGGATATACCAGTTATTGCCTATTCCTATTATGGCGATGAACAATAT 940

Qy         65 ArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnProLysValThrPhe 84
               |||    :::    |||||||||   |||      ||||||    |||
Db        941 CCTAGAACAATAAATATTCCATACCCAAAGGCTGGAGCTAAGAATCCCGTTGTTCGGATA 1000

Qy         85 LysMet---------------------------------SerGluIleMetIleAspAla 93
                 :::                                 |||:::   :::   |||
Db       1001 TTTATTATCGATACCACTTACCCTGCGTATGTAGGTCCCCAGGAAGTGCCTGTTCCAGCA 1060

Qy         94 ------------------------------------------------GluGlyArgIle 97
                                                              :::   |||:::
Db       1061 ATGATAGCCTCAAGTGATTATTATTTCAGTTGGCTCACGTGGGTTACTGATGAACGAGTA 1120

Qy         98 IleValAspGluValArgArgLeuValTyrPheGlu---------------------- 109
                 :::      ::::::|||:::
Db       1121 TGTTTGCAGTGGCTAAAAAGAGTCCAGAATGTTTCGGTCCTGTCTATATGTGACTTCAGG 1180

Qy        110 ---------GlyThrLysAspSerPro---LeuGluHisHisLeuTyr------------ 121
                      |||   |||   |||      :::   |||:::
Db       1181 GAAGACTGGCAGACATGGGATTGTCCAAAGACCCAGGAGCATATAGAAGAAAGCAGAACT 1240

Qy        122 --------------------------------ValValSerTyrValAsnProGlyGlu 130
                                             |||   ||||||
Db       1241 GGATGGGCTGGTGGATTCTTTGTTTCAAGACCAGTTTTCAGCTATGATGCCATTTCGTAC 1300

Qy        131 ValThrArgLeuThrAspArg---GlyTyrSerHisSerCysCysIleSerGlnHisCys 149
                  :::|||:::   ||||||   |||                :::||||||
Db       1301 TACAAAATATTTAGTGACAAGGATGGCTACAAACATAT-TCACTATATCAAAGACACTGT 1359

Qy        150 AspPhePheIleSerLysTyrSerAsnGln----------------------------- 159
                  |||   |||     |||
Db       1360 GGAAAATGCTATTCAAATTACAAGTGGCAAGTGGGAGGCCATAAATATATTCAGAGTAAC 1419

Qy        160 ------------------------------------------------LysAsnProHis 163
                                                           |||      |||
Db       1420 ACAGGATTCACTGTTTTATTCTAGCAATGAATTTGAAGAATACCCTGGAAGAAGAAACAT 1479

Qy        164 CysValSer---LeuTyrLysLeuSerSerProGluAsp-AspProThrCysLysThrLy 182
                  :::      :::||||||||||||   ::::::       ||||||      ::
Db       1480 CTACAGAATTAGCATTGGAAGCTATCCTCCAAGCAAGAAGTGTGTTACTTGCCATCTAAG 1539

Qy        182 sGlu---------------PheTrpAlaThrIleLeuAspSerAla-------------- 192
              ::::                 :::   |||:::      |||   |||
Db       1540 GAAAGAAAGGTGCCAATATTACACAGCAAGTTTCAGCGACTACGCCAAGTACTATGCACT 1599

Qy        193 ----------GlyPro---LeuProAspTyrThr------------------------- 199
                        ||||||   :::|||      |||
Db       1600 TGTCTGCTACGGCCCAGGCATCCCCATTTCCACCCTTCATGATGGACGCACTGATCAAGA 1659

Qy        200 ------------------------------------------------------------Pr 200
                                                                           ||
Db       1660 AATTAAAATCCTGGAAGAAAACAAGGAATTGGAAAATGCTTTGAAAAATATCCAGCTGCC 1719

Qy        200 oProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLy 220
              |   |||                          ||||||:::      :::
Db       1720 TAAAGAGGAAATTAAGAAACTTGAAGTAGATGAAATTACTTTATGGTACAAGATGATTCT 1779

Qy        220 sProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPr 240
               |||                 ||||||||||||   ::::::   :::||||||||||||
Db       1780 TCCTCCTCAATTTGACAGATCAAAGAAGTATCCCTTGCTAATTCAAGTGTATGGTGGTCC 1839

Qy        240 oGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLe 260
              |      |||   |||   :::                  |||   :::|||   ::
Db       1840 CTGCAGTCAGAGTGTAAGGTCT-----------------GTATTTGCTGTTAATTGGAT 1881

Qy        260 uAlaSerLeu-----------GlyTyrValValValValIleAspAsnArgGlySerCy 276
               ::::   |||          |||   |||:::   ::::::|||   ||||||:::
Db       1882 ATCTTATCTTGCAAGTAAGGAAGGGATGGTCATTGCCTTGGTGGATGGTCGAGGAACAGC 1941

Qy        276 sHisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAs 296
                 :::|||   |||          |||          |||:::|||       |||:::::
Db       1942 TTTCCAAGGTGACAAACTCCTCTATGCAGTGTATCGAAAGCTGGGTGTTTATGAAGTTGA 2001

Qy        296 pAspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspAr 316
               :||||||:::      ::::::                 ||||||||       ||
Db       2002 AGACCAGATTACAGCTGTCAGAAAATTCATA---GAAATGGGTTTCATTGATGAAAAAAG 2058

Qy        316 gValGlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnAr 336
               |:::   |||   |||||||||||||||||||||:::|||   :::|||||||
Db       2059 AATAGCCATATGGGGCTGGTCCTATGGAGGATACGTTTCATCACTGGCCCTTGCATCTGG 2118

Qy        336 gSerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAs 356
                :::   :::|||:::      ||||||   |||||||||:::   |||   :::|||
Db       2119 AACTGGTCTTTTCAAATGTGGTATAGCAGTGGCTCCAGTCTCCAGCTGGGAATATTACGC 2178

Qy        356 pThrGlyTyrThr 360
               :::   ||||||
Db       2179 GTCTGTCTACACA 2191

RESULT 8
US-08-940-391-1
; Sequence 1, Application US/08940391
; Patent No. 5965373
;  GENERAL INFORMATION:
;    APPLICANT:  Zimmermann, Rainer; Park, John E.;
;    APPLICANT:  Rettig, Wolfgang; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED DIMERIC FIBROBLAST ACTIVATION
;    TITLE OF INVENTION:  PROTEIN ALPHA, AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 2.0 MB  storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/940,391
```

```
;      FILING DATE:  01-OCT-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/619,280
;      FILING DATE:  18-MARCH-1996
;      APPLICATION NUMBER:  08/230,491
;      FILING DATE:  20-APRIL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5965373man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5330.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2815 Base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-940-391-1

Alignment Scores:
Pred. No.:                3.05e-20      Length:          2815
Score:                    257.00        Matches:         114
Percent Similarity:       36.70%        Conservative:    64
Best Local Similarity:    23.51%        Mismatches:      158
Query Match:              13.30%        Indels:          150
DB:                       2             Gaps:            20

US-10-070-464-7 (1-360) x US-08-940-391-1 (1-2815)

Qy          8 GlyValAlaThrPheValLeuGlnGluGluPhe------AspArgTyrSerGlyTyrTrp 25
              |||:::      :::|||   :::|||||||           :::|||:::      |||
Db        791 GGAATCCCAGACTGGGTTTATGAAGAGGAAATGCTTCCTACAAAATATGCT---CTCTGG 847

Qy         26 TrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeuArgIleLeuTyrGlu 45
              |||   |||                        ||||||   |||  .      |||
Db        848 TGGTCTCCT--------------------AATGGAAAATTTTTG------GCATATGCG 880

Qy         46 GluAsnAspGluSerGluValGluIleIleHisValThrSerProMetLeuGluThr--- 64
              |||   ::::::   ::::::   :::|||       :::              |||
Db        881 GAATTTAATGATAAGGATATACCAGTTATTGCCTATTCCTATTATGGCGATGAACAATAT 940

Qy         65 ArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnProLysValThrPhe 84
                |||      :::      |||||||||   |||      ||||||   |||
Db        941 CCTAGAACAATAAATATTCCATACCCAAAGGCTGGAGCTAAGAATCCCGTTGTTCGGATA 1000

Qy         85 LysMet---------------------------------SerGluIleMetIleAspAla 93
              :::                                   |||:::   :::   |||
Db       1001 TTTATTATCGATACCACTTACCCTGCGTATGTAGGTCCCCAGGAAGTGCCTGTTCCAGCA 1060

Qy         94 ------------------------------------------------GluGlyArgIle 97
                                                             :::   |||:::
Db       1061 ATGATAGCCTCAAGTGATTATTATTTCAGTTGGCTCACGTGGGTTACTGATGAACGAGTA 1120

Qy         98 IleValAspGluValArgArgLeuValTyrPheGlu---------------------- 109
              :::      ::::::|||:::
Db       1121 TGTTTGCAGTGGCTAAAAAGAGTCCAGAATGTTTCGGTCCTGTCTATATGTGACTTCAGG 1180

Qy        110 ---------GlyThrLysAspSerPro---LeuGluHisHisLeuTyr------------ 121
                       |||   |||   |||      :::   |||:::
Db       1181 GAAGACTGGCAGACATGGGATTGTCCAAAGACCCAGGAGCATATAGAAGAAAGCAGAACT 1240

Qy        122 --------------------------------ValValSerTyrValAsnProGlyGlu 130
                                             |||   ||||||
Db       1241 GGATGGGCTGGTGGATTCTTTGTTTCAAGACCAGTTTTCAGCTATGATGCCATTTCGTAC 1300

Qy        131 ValThrArgLeuThrAspArg---GlyTyrSerHisSerCysCysIleSerGlnHisCys 149
                  :::|||:::   ||||||   |||           :::||||||
Db       1301 TACAAAATATTTAGTGACAAGGATGGCTACAAACATAT-TCACTATATCAAAGACACTGT 1359

Qy        150 AspPhePheIleSerLysTyrSerAsnGln------------------------------ 159
```

```
                        |||   |||      |||
Db       1360 GGAAAATGCTATTCAAATTACAAGTGGCAAGTGGGAGGCCATAAATATATTCAGAGTAAC 1419

Qy        160 ----------------------------------------------LysAsnProHis 163
                                                           |||      |||
Db       1420 ACAGGATTCACTGTTTTATTCTAGCAATGAATTTGAAGAATACCCTGGAAGAAGAAACAT 1479

Qy        164 CysValSer---LeuTyrLysLeuSerSerProGluAsp-AspProThrCysLysThrLy 182
                  :::      :::|||||||||||    ::::::      |||||     ::
Db       1480 CTACAGAATTAGCATTGGAAGCTATCCTCCAAGCAAGAAGTGTGTTACTTGCCATCTAAG 1539

Qy        182 sGlu---------------PheTrpAlaThrIleLeuAspSerAla-------------- 192
              ::::               :::   |||:::      |||   |||
Db       1540 GAAAGAAAGGTGCCAATATTACACAGCAAGTTTCAGCGACTACGCCAAGTACTATGCACT 1599

Qy        193 ----------GlyPro---LeuProAspTyrThr------------------------- 199
                      ||||||   :::|||      |||
Db       1600 TGTCTGCTACGGCCCAGGCATCCCCATTTCCACCCTTCATGATGGACGCACTGATCAAGA 1659

Qy        200 ----------------------------------------------------------Pr 200
                                                                         ||
Db       1660 AATTAAAATCCTGGAAGAAAACAAGGAATTGGAAAATGCTTTGAAAAATATCCAGCTGCC 1719

Qy        200 oProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLy 220
              |   |||                                 ||||||:::      :::
Db       1720 TAAAGAGGAAATTAAGAAACTTGAAGTAGATGAAATTACTTTATGGTACAAGATGATTCT 1779

Qy        220 sProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPr 240
              |||                  ||||||||||||   ::::::   :::||||||||||||
Db       1780 TCCTCCTCAATTTGACAGATCAAAGAAGTATCCCTTGCTAATTCAAGTGTATGGTGGTCC 1839

Qy        240 oGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLe 260
              |       |||   |||   :::                    |||  :::|||   ::
Db       1840 CTGCAGTCAGAGTGTAAGGTCT-----------------GTATTTGCTGTTAATTGGAT 1881

Qy        260 uAlaSerLeu------------GlyTyrValValValValIleAspAsnArgGlySerCy 276
              ::::   |||            |||   |||:::   ::::::|||   ||||||:::
Db       1882 ATCTTATCTTGCAAGTAAGGAAGGGATGGTCATTGCCTTGGTGGATGGTCGAGGAACAGC 1941

Qy        276 sHisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAs 296
                 :::|||   |||          |||          |||:::|||      |||:::::
Db       1942 TTTCCAAGGTGACAAACTCCTCTATGCAGTGTATCGAAAGCTGGGTGTTTATGAAGTTGA 2001

Qy        296 pAspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspAr 316
              :||||||:::      ::::::                  |||||||||      ||
Db       2002 AGACCAGATTACAGCTGTCAGAAAATTCATA---GAAATGGGTTTCATTGATGAAAAAAG 2058

Qy        316 gValGlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnAr 336
              |:::   |||   ||||||||||||||||||||:::|||   :::|||||||
Db       2059 AATAGCCATATGGGGCTGGTCCTATGGAGGATACGTTTCATCACTGGCCCTTGCATCTGG 2118

Qy        336 gSerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAs 356
               :::   :::|||:::      ||||||   |||||||||:::   |||   :::|||
Db       2119 AACTGGTCTTTTCAAATGTGGTATAGCAGTGGCTCCAGTCTCCAGCTGGGAATATTACGC 2178

Qy        356 pThrGlyTyrThr 360
              :::   ||||||
Db       2179 GTCTGTCTACACA 2191

RESULT 9
US-09-221-017B-253
; Sequence 253, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
```

```
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  253:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  543 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1...543
US-09-221-017B-253

Alignment Scores:
Pred. No.:              6.28e-19        Length:         543
Score:                  236.50          Matches:        41
Percent Similarity:     65.00%          Conservative:   24
Best Local Similarity:  41.00%          Mismatches:     34
Query Match:            12.23%          Indels:         1
DB:                     4               Gaps:           1

US-10-070-464-7 (1-360) x US-09-221-017B-253 (1-543)

Qy        260 LeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCysHisArgGly 279
              :::|||      ||||||   |||      :::|||:::|||||||||   :::|||||
Db         43 ATGGCACAGAAAGGCTATGCCGTCTTTACGGTGGATAGTCGCGGATCTGCCAATAGAGGG 102

Qy        280 LeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnVal 299
                  ||||||               ::::::||||||   |||:::   ||||||:::
Db        103 GCTGCTTTCGAGCAGGTTATTCATCGTCGTTTGGGGCAGACCGAGATGGCCGATCAGATG 162

Qy        300 GluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIle 319
              |||:::   :::|||   |||:::      :::::::|||   ||||||:::|||:::
Db        163 TGCGGTGTGGATTTCCTCAAGAGCCAA---TCATGGGTGGATGCCGATAGAATAGGAGTA 219

Qy        320 HisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIle 339
              ||||||||||||||||||||:::::::::         ::::::         |||:::
Db        220 CATGGCTGGAGCTATGGTGGCTTTATGACTACGAATCTGATGCTTACGCACGGCGATGTC 279

Qy        340 PheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyr 359
              |||:::|||   :::||||||   ||||||      |||      |||:::      |||
Db        280 TTCAAAGTCGGAGTAGCCGGCGGGCCTGTCATAGACTGGAATCGATATGAGATTATGTAC 339

RESULT 10
US-09-389-681-428
; Sequence 428, Application US/09389681A
; Patent No. 6518237
; GENERAL INFORMATION:
;  APPLICANT: Yuqui, Jiang
;  APPLICANT: Dillon, Davin C.
;  APPLICANT: Mitcham, Jennifer L.
;  APPLICANT: Xu, Jiangchun
;  TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
;  TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.470C3
;  CURRENT APPLICATION NUMBER: US/09/389,681A
;  CURRENT FILING DATE: 1999-09-02
;  NUMBER OF SEQ ID NOS: 463
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 428
;   LENGTH: 535
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-389-681-428

Alignment Scores:
Pred. No.:              1.4e-18         Length:         535
Score:                  233.50          Matches:        52
Percent Similarity:     54.86%          Conservative:   27
Best Local Similarity:  36.11%          Mismatches:     54
Query Match:            12.08%          Indels:         11
DB:                     4               Gaps:           3

US-10-070-464-7 (1-360) x US-09-389-681-428 (1-535)

Qy        221 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPro 240
              |||                  ||||||||||||   ::::::   :::|||||||||||||
Db         15 CCTCCTCAATTTGACAGATCAAAGAAGTATCCCTTGCTAATTCAAGTGTATGGTGGTCCC 74

Qy        241 GlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeu 260
                    |||   |||   :::                  |||   :::|||   :::
Db         75 TGCAGTCAGAGTGTAAGGTCT-----------------GTTTTGCTGTTAATTGGATA 116

Qy        261 AlaSerLeu-----------GlyTyrValValValValIleAspAsnArgGlySerCys 276
              :::   |||           |||   |||:::   :::::::|||   ||||||:::
Db        117 TCTTATCTTGCAAGTAAGGAAGGGATGGTCATTGCCTTGGTGGATGGTCGAGGAACAGCT 176

Qy        277 HisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAsp 296
                 :::|||   |||         |||         |||:::|||      |||::::::
Db        177 TTCCAAGGTGACAAACTCCTCTATGCAGTGTATCGAAAGCTGGGTGTTTATGAAGTTGAA 236

Qy        297 AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 316
              ||||||:::      ::::::                  |||||||||      |||
Db        237 GACCAGATTACAGCTGTCAGAAAATTCATA---GAAATGGGTTTCATTGATGAAAAAAGA 293

Qy        317 ValGlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 336
              :::   |||   ||||||||||||||||||||:::|||   :::||||||
Db        294 ATAGCCATATGGGGCTGGTCCTATGGAGGATACGTTTCATCACTGGCCCTTGCATCTGGA 353

Qy        337 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAsp 356
              :::   :::|||:::      ||||||   |||||||||:::   |||   :::|||
Db        354 ACTGGTCTTTTCAAATGTGGTATAGCAGTGGCTCCAGTCTCCAGCTGGGAATATTACGCG 413

Qy        357 ThrGlyTyrThr 360
              :::   ||||||
Db        414 TCTGTCTACACA 425

RESULT 11
US-09-620-405B-428
; Sequence 428, Application US/09620405B
; Patent No. 6528054
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Yuqiu
```

```
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Hepler, William T.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF BREAST CANCER
;  FILE REFERENCE: 210121.470C8
;  CURRENT APPLICATION NUMBER: US/09/620,405B
;  CURRENT FILING DATE:  2000-07-20
;  NUMBER OF SEQ ID NOS: 495
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 428
;   LENGTH: 535
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-620-405B-428

Alignment Scores:
Pred. No.:              1.4e-18       Length:          535
Score:                  233.50        Matches:         52
Percent Similarity:     54.86%        Conservative:    27
Best Local Similarity:  36.11%        Mismatches:      54
Query Match:            12.08%        Indels:          11
DB:                     4             Gaps:            3

US-10-070-464-7 (1-360) x US-09-620-405B-428 (1-535)

Qy        221 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPro 240
              |||             |||||||||||   ::::::   :::|||||||||||||
Db         15 CCTCCTCAATTTGACAGATCAAAGAAGTATCCCTTGCTAATTCAAGTGTATGGTGGTCCC 74

Qy        241 GlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeu 260
              |||   |||   :::                  |||   :::|||   :::
Db         75 TGCAGTCAGAGTGTAAGGTCT----------------GTATTTGCTGTTAATTGGATA 116

Qy        261 AlaSerLeu-----------GlyTyrValValValValIleAspAsnArgGlySerCys 276
              :::   |||             |||   |||:::   ::::::|||   ||||||:::
Db        117 TCTTATCTTGCAAGTAAGGAAGGGATGGTCATTGCCTTGGTGGATGGTCGAGGAACAGCT 176

Qy        277 HisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAsp 296
              :::|||   |||         |||         |||:::|||       |||::::::
Db        177 TTCCAAGGTGACAAACTCCTCTATGCAGTGTATCGAAAGCTGGGTGTTTATGAAGTTGAA 236

Qy        297 AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 316
              ||||||:::     ::::::                  |||||||||      |||
Db        237 GACCAGATTACAGCTGTCAGAAAATTCATA---GAAATGGGTTTCATTGATGAAAAAAGA 293

Qy        317 ValGlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 336
              :::   |||   |||||||||||||||||||:::|||   :::||||||
Db        294 ATAGCCATATGGGGCTGGTCCTATGGAGGATACGTTTCATCACTGGCCCTTGCATCTGGA 353

Qy        337 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAsp 356
              :::   :::|||:::      ||||||   ||||||||||:::   |||   :::|||
Db        354 ACTGGTCTTTTCAAATGTGGTATAGCAGTGGCTCCAGTCTCCAGCTGGGAATATTACGCG 413

Qy        357 ThrGlyTyrThr 360
              :::   ||||||
Db        414 TCTGTCTACACA 425

RESULT 12
US-09-433-826B-428
; Sequence 428, Application US/09433826B
; Patent No. 6579973
; GENERAL INFORMATION:
;  APPLICANT:  Jiang, Yuqui
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Harlocker, Susan L.
;  TITLE OF INVENTION:  COMPOSITIONS FOR THE TREATMENT AND
;  TITLE OF INVENTION:  DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE:  210121.470C4
;  CURRENT APPLICATION NUMBER: US/09/433,826B
;  CURRENT FILING DATE:  1999-11-03
;  NUMBER OF SEQ ID NOS:  474
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 428
;   LENGTH: 535
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-433-826B-428

Alignment Scores:
Pred. No.:              1.4e-18       Length:          535
Score:                  233.50        Matches:         52
Percent Similarity:     54.86%        Conservative:    27
Best Local Similarity:  36.11%        Mismatches:      54
Query Match:            12.08%        Indels:          11
DB:                     4             Gaps:            3

US-10-070-464-7 (1-360) x US-09-433-826B-428 (1-535)

Qy        221 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPro 240
              |||             |||||||||||   ::::::   :::|||||||||||||
Db         15 CCTCCTCAATTTGACAGATCAAAGAAGTATCCCTTGCTAATTCAAGTGTATGGTGGTCCC 74

Qy        241 GlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeu 260
              |||   |||   :::                  |||   :::|||   :::
Db         75 TGCAGTCAGAGTGTAAGGTCT----------------GTATTTGCTGTTAATTGGATA 116

Qy        261 AlaSerLeu-----------GlyTyrValValValValIleAspAsnArgGlySerCys 276
              :::   |||             |||   |||:::   ::::::|||   ||||||:::
Db        117 TCTTATCTTGCAAGTAAGGAAGGGATGGTCATTGCCTTGGTGGATGGTCGAGGAACAGCT 176

Qy        277 HisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAsp 296
              :::|||   |||         |||         |||:::|||       |||::::::
Db        177 TTCCAAGGTGACAAACTCCTCTATGCAGTGTATCGAAAGCTGGGTGTTTATGAAGTTGAA 236

Qy        297 AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 316
              ||||||:::     ::::::                  |||||||||      |||
Db        237 GACCAGATTACAGCTGTCAGAAAATTCATA---GAAATGGGTTTCATTGATGAAAAAAGA 293

Qy        317 ValGlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 336
              :::   |||   |||||||||||||||||||:::|||   :::||||||
Db        294 ATAGCCATATGGGGCTGGTCCTATGGAGGATACGTTTCATCACTGGCCCTTGCATCTGGA 353

Qy        337 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAsp 356
              :::   :::|||:::      ||||||   ||||||||||:::   |||   :::|||
Db        354 ACTGGTCTTTTCAAATGTGGTATAGCAGTGGCTCCAGTCTCCAGCTGGGAATATTACGCG 413

Qy        357 ThrGlyTyrThr 360
              :::   ||||||
Db        414 TCTGTCTACACA 425

RESULT 13
US-09-604-287A-428
; Sequence 428, Application US/09604287A
; Patent No. 6586572
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Yuqiu
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Hepler, William T.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF BREAST CANCER
;  FILE REFERENCE: 210121.470C7
;  CURRENT APPLICATION NUMBER: US/09/604,287A
;  CURRENT FILING DATE:  2000-06-22
;  NUMBER OF SEQ ID NOS: 489
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 428
;   LENGTH: 535
;   TYPE: DNA
```

```
;  ORGANISM: Homo sapiens
US-09-604-287A-428

Alignment Scores:
Pred. No.:              1.4e-18       Length:          535
Score:                  233.50        Matches:         52
Percent Similarity:     54.86%        Conservative:    27
Best Local Similarity:  36.11%        Mismatches:      54
Query Match:            12.08%        Indels:          11
DB:                     4             Gaps:            3

US-10-070-464-7 (1-360) x US-09-604-287A-428 (1-535)

Qy        221 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyPro 240
              |||                 ||||||||||||   ::::::   :::||||||||||||
Db         15 CCTCCTCAATTTGACAGATCAAAGAAGTATCCCTTGCTAATTCAAGTGTATGGTGGTCCC 74

Qy        241 GlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeu 260
                  |||   |||   :::                     |||   :::|||   :::
Db         75 TGCAGTCAGAGTGTAAGGTCT-----------------GTATTTGCTGTTAATTGGATA 116

Qy        261 AlaSerLeu-----------GlyTyrValValValValIleAspAsnArgGlySerCys 276
              :::   |||           |||   |||:::   ::::::|||   ||||||:::
Db        117 TCTTATCTTGCAAGTAAGGAAGGGATGGTCATTGCCTTGGTGGATGGTCGAGGAACAGCT 176

Qy        277 HisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAsp 296
                ::::|||   |||         |||         |||:::|||         |||::::::
Db        177 TTCCAAGGTGACAAACTCCTCTATGCAGTGTATCGAAAGCTGGGTGTTTATGAAGTTGAA 236

Qy        297 AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 316
              ||||||:::     ::::::                    |||||||||      |||
Db        237 GACCAGATTACAGCTGTCAGAAAATTCATA---GAAATGGGTTTCATTGATGAAAAAAGA 293

Qy        317 ValGlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 336
              :::   |||   |||||||||||||||||||||:::|||   :::|||||||
Db        294 ATAGCCATATGGGGCTGGTCCTATGGAGGATACGTTTCATCACTGGCCCTTGCATCTGGA 353

Qy        337 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAsp 356
              :::   :::|||:::     ||||||  |||||||||:::   |||   :::|||
Db        354 ACTGGTCTTTTCAAATGTGGTATAGCAGTGGCTCCAGTCTCCAGCTGGGAATATTACGCG 413

Qy        357 ThrGlyTyrThr 360
              :::   ||||||
Db        414 TCTGTCTACACA 425

RESULT 14
US-08-699-103B-1/c
; Sequence 1, Application US/08699103B
; Patent No. 6107462
;  GENERAL INFORMATION:
;    APPLICANT:  Rine, Jasper D.
;    APPLICANT:  Hampton, Randolph
;    TITLE OF INVENTION:  GENES AND PROTEINS CONTROLLING
;    TITLE OF INVENTION:  CHOLESTEROL SYNTHESIS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  2200 Sand Hill Road, Suite 100
;      CITY:  Menlo Park
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/699,103B
;      FILING DATE:  16-AUG-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/002,581
;      FILING DATE:  17-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Green, Grant D.
;      REGISTRATION NUMBER:  31,259
;      REFERENCE/DOCKET NUMBER:  09272/005001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/322-5070
;      TELEFAX:  650/854-0875
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4982 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-699-103B-1

Alignment Scores:
Pred. No.:              2.26e-14      Length:          4982
Score:                  211.00        Matches:         49
Percent Similarity:     50.74%        Conservative:    20
Best Local Similarity:  36.03%        Mismatches:      55
Query Match:            10.92%        Indels:          12
DB:                     3             Gaps:            4

US-10-070-464-7 (1-360) x US-08-699-103B-1 (1-4982)

Qy        230 TyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsnArgPhe 249
              ||||||          |||  ||||||||||||      |||   |||
Db       4941 TATCCTGTATTTTTCTTTGCATATGGGGGACCGAATTCTCAACAAGTT------------ 4894

Qy        250 LysGlyValLysTyrPheArgLeuAsn-----------ThrLeuAlaSerLeuGlyTyr 265
                  ||||||   |||   :::                         :::   |||
Db       4893 ------GTCAAAACGTTTTCCGTAGGATTTAATGAAGTGGTAGCTTCACAATTAAACGCA 4840

Qy        266 ValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAla 285
              :::||||||||||:::|||   ||||||:::     :::|||     |||
Db       4839 ATTGTAGTTGTTGTTGACGGTCGTGGTACTGGCTTCAAAGGTCAAGACTTTAGATCCCTT 4780

Qy        286 PheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeu 305
                 :::   ::::::|||      |||      ||||||:::
Db       4779 GTTCGCGATAGGCTCGGTGATTACGAGGCCCGCGACCAAATATCTGCGGCTTCCTTATAT 4720

Qy        306 AlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGly 325
              |||          |||:::|||      ::::::   :::   ||||||||||||||
Db       4719 GGTTCT---TTAACTTTTGTTGATCCGCAAAAGATTTCCTTATTTGGTTGGTCATACGGG 4663

Qy        326 GlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAsp---IlePheArgValAlaIle 344
              |||||||||:::|||      |||   :::          |||:::      :::
Db       4662 GGGTACCTGACACTAAAAACTTTGGAGAAAGATGGCGGAAGACATTTCAAATACGGGATG 4603

Qy        345 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              :::   ||||||||||||   |||   |||||||||:::   ||||||
Db       4602 TCAGTTGCGCCAGTAACCGACTGGAGATTTTACGATTCTGTTTATACT 4555

RESULT 15
US-09-229-059-1/c
; Sequence 1, Application US/09229059
; Patent No. 6333172
;  GENERAL INFORMATION:
;    APPLICANT:  Rine, Jasper D.
;    APPLICANT:  Hampton, Randolph
;    TITLE OF INVENTION:  GENES AND PROTEINS CONTROLLING
;    TITLE OF INVENTION:  CHOLESTEROL SYNTHESIS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  2200 Sand Hill Road, Suite 100
;      CITY:  Menlo Park
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
```

```
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/229,059
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/699,103
;      FILING DATE:  16-AUG-1996
;      APPLICATION NUMBER:  60/002,581
;      FILING DATE:  17-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Green, Grant D.
;      REGISTRATION NUMBER:  31,259
;      REFERENCE/DOCKET NUMBER:  09272/005001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/322-5070
;      TELEFAX:  650/854-0875
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4982 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-09-229-059-1

Alignment Scores:
Pred. No.:                2.26e-14      Length:          4982
Score:                    211.00        Matches:         49
Percent Similarity:       50.74%        Conservative:    20
Best Local Similarity:    36.03%        Mismatches:      55
Query Match:              10.92%        Indels:          12
DB:                       4             Gaps:            4

US-10-070-464-7 (1-360) x US-09-229-059-1 (1-4982)

Qy        230 TyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsnArgPhe 249
              ||||||          |||   ||||||||||||       |||   |||
Db       4941 TATCCTGTATTTTTCTTTGCATATGGGGGACCGAATTCTCAACAAGTT------------ 4894

Qy        250 LysGlyValLysTyrPheArgLeuAsn------------ThrLeuAlaSerLeuGlyTyr 265
                  ||||||   |||   :::                    :::   |||
Db       4893 ------GTCAAAACGTTTTCCGTAGGATTTAATGAAGTGGTAGCTTCACAATTAAACGCA 4840

Qy        266 ValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAla 285
              :::||||||||||:::|||   ||||||:::      :::|||      |||
Db       4839 ATTGTAGTTGTTGTTGACGGTCGTGGTACTGGCTTCAAAGGTCAAGACTTTAGATCCCTT 4780

Qy        286 PheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeu 305
                 :::   ::::::|||       |||       ||||||:::
Db       4779 GTTCGCGATAGGCTCGGTGATTACGAGGCCCGCGACCAAATATCTGCGGCTTCCTTATAT 4720

Qy        306 AlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGly 325
                 |||          |||:::|||      ::::::   :::   ||||||||||||||
Db       4719 GGTTCT---TTAACTTTTGTTGATCCGCAAAAGATTTCCTTATTTGGTTGGTCATACGGG 4663

Qy        326 GlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAsp---IlePheArgValAlaIle 344
              |||||||||:::|||      |||   :::          |||:::      :::
Db       4662 GGGTACCTGACACTAAAAACTTTGGAGAAAGATGGCGGAAGACATTTCAAATACGGGATG 4603

Qy        345 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              :::   ||||||||||||   |||   |||||||||:::   ||||||
Db       4602 TCAGTTGCGCCAGTAACCGACTGGAGATTTTACGATTCTGTTTATACT 4555


Search completed: October 16, 2003, 03:31:25
Job time : 81.4006 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 16, 2003, 00:05:32 ; Search time 269.152 Seconds
(without alignments)
3512.533 Million cell updates/sec

Title: US-10-070-464-7
Perfect score: 1933
Sequence: 1 EEDARSAGVATFVLQEEFDR.....RVAIAGAPVTLWIFYDTGYT 360

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1750203 seqs, 1313063994 residues

Total number of hits satisfying chosen parameters: 3500406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10070464/runat_15102003_113555_24924/app_query.fasta_1.2652
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10070464_@CGN_1_1_845_@runat_15102003_113555_24924
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1880	97.3	4309	10	US-09-976-674-14	Sequence 14, Appl
2	1836.5	95.0	2649	12	US-10-054-776-1	Sequence 1, Appli
3	1836.5	95.0	2649	12	US-10-170-789-39	Sequence 39, Appl
4	1836.5	95.0	2671	10	US-09-976-674-2	Sequence 2, Appli
5	1836.5	95.0	3143	12	US-10-170-789-37	Sequence 37, Appl
6	1816.5	94.0	4829	10	US-09-976-674-12	Sequence 12, Appl
7	1522	78.7	4685	10	US-09-976-674-22	Sequence 22, Appl
8	1522.5	78.7	4676	10	US-09-976-674-20	Sequence 20, Appl
9	1229	63.6	4523	10	US-09-976-674-8	Sequence 8, Appli
10	1220.5	63.1	2617	10	US-09-976-674-4	Sequence 4, Appli
11	1220.5	63.1	4219	10	US-09-976-674-28	Sequence 28, Appl
12	1220.5	63.1	4302	10	US-09-976-674-24	Sequence 24, Appl
13	1201	62.1	4180	10	US-09-976-674-36	Sequence 36, Appl
14	1201	62.1	4263	10	US-09-976-674-34	Sequence 34, Appl
15	1183.5	61.2	3262	13	US-10-098-841-83	Sequence 83, Appl
16	1169.5	60.5	2801	13	US-10-098-841-100	Sequence 100, App
17	1132.5	58.6	4076	10	US-09-976-674-32	Sequence 32, Appl
18	1132.5	58.6	4159	10	US-09-976-674-30	Sequence 30, Appl
19	1113	57.6	4037	10	US-09-976-674-40	Sequence 40, Appl
20	1113	57.6	4120	10	US-09-976-674-38	Sequence 38, Appl
21	630	32.6	502	11	US-09-918-995-19585	Sequence 19585, A
22	368	19.0	561	11	US-09-764-891-877	Sequence 877, App
23	352	18.2	2079	13	US-10-044-090-843	Sequence 843, App
24	352	18.2	2411	10	US-09-976-674-26	Sequence 26, Appl
25	315	16.3	3407	12	US-10-423-714-5	Sequence 5, Appli
26	315	16.3	3407	13	US-10-002-593-5	Sequence 5, Appli
27	315	16.3	3407	14	US-10-165-603-6	Sequence 6, Appli
28	308.5	16.0	2130	14	US-10-156-761-3131	Sequence 3131, Ap
29	308.5	16.0	9025608	14	US-10-156-761-1	Sequence 1, Appli
30	297.5	15.4	4835	10	US-09-917-800A-1570	Sequence 1570, Ap
31	297.5	15.4	4835	14	US-10-165-603-5	Sequence 5, Appli
32	271	14.0	4797	11	US-09-764-891-7074	Sequence 7074, Ap
33	261	13.5	2812	14	US-10-084-817-24	Sequence 24, Appl
34	258.5	13.4	2366	12	US-10-101-510-683	Sequence 683, App
35	258.5	13.4	2788	12	US-10-269-909-22	Sequence 22, Appl
36	258.5	13.4	3138	14	US-10-198-846-13171	Sequence 13171, A
37	257.5	13.3	3224	12	US-10-240-965-117	Sequence 117, App
38	257	13.3	2814	10	US-09-962-832-108	Sequence 108, App
39	257	13.3	2814	10	US-09-954-456-1148	Sequence 1148, Ap
40	257	13.3	2814	10	US-09-954-531-367	Sequence 367, App
41	257	13.3	2814	12	US-10-101-510-30	Sequence 30, Appl
42	257	13.3	2814	12	US-10-301-822-54	Sequence 54, Appl
43	257	13.3	2814	12	US-09-873-367C-499	Sequence 499, App
44	257	13.3	2814	14	US-10-177-293-135	Sequence 135, App
45	257	13.3	2815	9	US-09-265-606-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-976-674-14
; Sequence 14, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 4309
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-14

Alignment Scores:
Pred. No.: 5.23e-231 Length: 4309

Score:	1880.00	Matches:	358
Percent Similarity:	98.90%	Conservative:	0
Best Local Similarity:	98.90%	Mismatches:	2
Query Match:	97.26%	Indels:	3
DB:	10	Gaps:	0

US-10-070-464-7 (1-360) x US-09-976-674-14 (1-4309)

```
Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1002

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1062

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1122

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1182

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIleValAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGATCCAA-GTTGAT 1241

Qy        101 GluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1242 GAAGTCAGAAGGCTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTG 1301

Qy        121 TyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1302 TACGTAGTCAGTTACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCA 1361

Qy        141 HisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLys 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1362 CATTCTTGCTGCATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAG 1421

Qy        161 AsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLys 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1422 AATCCACACTGTGTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAA 1481

Qy        181 ThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1482 ACAAAGGAATTTTGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCT 1541

Qy        201 ProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLys 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1542 CCAGAAATTTTCTCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAG 1601

Qy        221 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly--P 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  | |
Db       1602 CCTCATGATCTACAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCTC 1661

Qy        240 roGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrL 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1662 CTCAGGTGCAGTTGGTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCC 1721

Qy        260 euAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCysHisArgGlyL 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1722 TAGCCTCTCTAGGTTATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGC 1781

Qy        280 euLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1782 TTAAATTTGAAGGCGCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGG 1841

Qy        300 luGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleH 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1842 AAGGACTCCAATATCTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCC 1901

Qy        320 isGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIleP 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1902 ACGGCTGGTCCTATGGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCT 1961

Qy        340 heArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1962 TCAGGGTTGCTATTGCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACA 2021

Qy        360 hr 360
              ||
Db       2022 CG 2023
```

RESULT 2
US-10-054-776-1
; Sequence 1, Application US/10054776
; Publication No. US20030165818A1
; GENERAL INFORMATION:
; APPLICANT: Mark Robert Edbrooke
; APPLICANT: Alan Peter Lewis
; TITLE OF INVENTION: NOVEL PROTEIN
; FILE REFERENCE: QG1042US
; CURRENT APPLICATION NUMBER: US/10/054,776
; CURRENT FILING DATE: 2002-01-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2649
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2649)
US-10-054-776-1

Alignment Scores:			
Pred. No.:	9.81e-226	Length:	2649
Score:	1836.50	Matches:	360
Percent Similarity:	67.54%	Conservative:	0
Best Local Similarity:	67.54%	Mismatches:	0
Query Match:	95.01%	Indels:	173
DB:	12	Gaps:	1

US-10-070-464-7 (1-360) x US-10-054-776-1 (1-2649)

```
Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        730 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 789

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        790 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 849

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        850 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 909

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        910 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 969

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        970 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1029

Qy         98 ---------------------------------------------------------- 98

Db       1030 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1089

Qy         98 ---------------------------------------------------------- 98

Db       1090 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1149

Qy         98 ---------------------------------------------------------- 98
```

```
Db       1150 ACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1209

Qy         98 ------------------------------------------------------------ 98

Db       1210 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1269

Qy         98 ------------------------------------------------------------ 98

Db       1270 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1329

Qy         98 ------------------------------------------------------------ 98

Db       1330 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1389

Qy         98 ------------------------------------------------------------ 98

Db       1390 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1449

Qy         98 ------------------------------------------------------------ 98

Db       1450 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1509

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                 ||||||||||||||||||||||||
Db       1510 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1569

Qy       108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1570 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1629

Qy       128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1630 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1689

Qy       148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1690 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1749

Qy       168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1750 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 1809

Qy       188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1810 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 1869

Qy       208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1870 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 1929

Qy       228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1930 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 1989

Qy       248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1990 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2049

Qy       268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2050 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2109

Qy       288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2110 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2169

Qy       308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2170 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2229

Qy       328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2230 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2289
```

```
Qy       348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
             ||||||||||||||||||||||||||||||||||||||
Db       2290 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2328
```

RESULT 3
US-10-170-789-39
; Sequence 39, Application US/10170789
; Publication No. US20030180930A1
; GENERAL INFORMATION:
; APPLICANT: Rachel E. Meyers
; APPLICANT: Olandt, Peter J.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Curtis , Rory A. J.
; APPLICANT: Williamson, Mark
; APPLICANT: Weich, Nadine
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,
; TITLE OF INVENTION: AND PROTEASE FAMILY MEMBERS AND USES THEREOF
; FILE REFERENCE: 10448-191001
; CURRENT APPLICATION NUMBER: US/10/170,789
; CURRENT FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 09/797,039
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/06525
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/186,061
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 09/882,166
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19269
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/212,078
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/934,406
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/26052
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/226,740
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/861,801
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: PCT/US01/16549
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 60/205,508
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/801,267
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07138
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,454
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/829,671
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: PCT/US01/40483
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: US 60/197,508
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/961,721
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/29904
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 60/235,023
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 10/045,367
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/246,561
; PRIOR FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: US 09/801,275
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: PCT/US01/07074
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 60/187,420
; PRIOR FILING DATE: 2000-03-07

```
;  NUMBER OF SEQ ID NOS: 63
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 39
;   LENGTH: 2649
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-170-789-39

Alignment Scores:
Pred. No.:              9.81e-226       Length:         2649
Score:                  1836.50         Matches:        360
Percent Similarity:     67.54%          Conservative:   0
Best Local Similarity:  67.54%          Mismatches:     0
Query Match:            95.01%          Indels:         173
DB:                     12              Gaps:           1

US-10-070-464-7 (1-360) x US-10-170-789-39 (1-2649)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        730 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 789

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        790 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 849

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        850 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 909

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        910 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 969

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        970 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1029

Qy         98 ---------------------------------------------------------- 98

Db       1030 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1089

Qy         98 ---------------------------------------------------------- 98

Db       1090 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1149

Qy         98 ---------------------------------------------------------- 98

Db       1150 ACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1209

Qy         98 ---------------------------------------------------------- 98

Db       1210 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1269

Qy         98 ---------------------------------------------------------- 98

Db       1270 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1329

Qy         98 ---------------------------------------------------------- 98

Db       1330 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1389

Qy         98 ---------------------------------------------------------- 98

Db       1390 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1449

Qy         98 ---------------------------------------------------------- 98

Db       1450 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1509

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                 |||||||||||||||||||||||
Db       1510 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1569

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1570 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1629

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1630 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1689

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1690 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1749

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1750 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 1809

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1810 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 1869

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1870 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 1929

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1930 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 1989

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1990 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2049

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2050 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2109

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2110 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2169

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2170 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2229

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2230 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2289

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||||
Db       2290 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2328

RESULT 4
US-09-976-674-2
; Sequence 2, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 2671
;   TYPE: DNA
```

```
;   ORGANISM: Homo sapiens
US-09-976-674-2

Alignment Scores:
Pred. No.:              9.94e-226       Length:         2671
Score:                  1836.50         Matches:        360
Percent Similarity:     67.54%          Conservative:   0
Best Local Similarity:  67.54%          Mismatches:     0
Query Match:            95.01%          Indels:         173
DB:                     10              Gaps:           1

US-10-070-464-7 (1-360) x US-09-976-674-2 (1-2671)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        737 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 796

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        797 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 856

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        857 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 916

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        917 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 976

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1036

Qy         98 ---------------------------------------------------------- 98

Db       1037 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1096

Qy         98 ---------------------------------------------------------- 98

Db       1097 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1156

Qy         98 ---------------------------------------------------------- 98

Db       1157 ACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1216

Qy         98 ---------------------------------------------------------- 98

Db       1217 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1276

Qy         98 ---------------------------------------------------------- 98

Db       1277 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1336

Qy         98 ---------------------------------------------------------- 98

Db       1337 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1396

Qy         98 ---------------------------------------------------------- 98

Db       1397 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1456

Qy         98 ---------------------------------------------------------- 98

Db       1457 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1516

Qy         99 ------------------------------ValAspGluValArgArgLeuValTyr 107
                                            |||||||||||||||||||||||||
Db       1517 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1576

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1577 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1636

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1637 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1696

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1697 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1756

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1757 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 1816

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1817 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 1876

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1877 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 1936

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1937 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 1996

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1997 CGATTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2056

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2057 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2116

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2117 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2176

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2177 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2236

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2237 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2296

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||||
Db       2297 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2335

RESULT 5
US-10-170-789-37
; Sequence 37, Application US/10170789
; Publication No. US20030180930A1
; GENERAL INFORMATION:
;  APPLICANT: Rachel E. Meyers
;  APPLICANT:  Olandt, Peter J.
;  APPLICANT:  Kapeller-Libermann, Rosana
;  APPLICANT:  Curtis , Rory A. J.
;  APPLICANT:  Williamson, Mark
;  APPLICANT:  Weich, Nadine
;  TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,
;  TITLE OF INVENTION:  AND PROTEASE FAMILY MEMBERS AND USES THEREOF
;  FILE REFERENCE: 10448-191001
;  CURRENT APPLICATION NUMBER: US/10/170,789
;  CURRENT FILING DATE:  2002-06-13
;  PRIOR APPLICATION NUMBER: US 09/797,039
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: PCT/US01/06525
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: US 60/186,061
;  PRIOR FILING DATE: 2000-02-29
;  PRIOR APPLICATION NUMBER: US 09/882,166
;  PRIOR FILING DATE: 2001-06-15
;  PRIOR APPLICATION NUMBER: PCT/US01/19269
```

```
;  PRIOR FILING DATE: 2001-06-15
;  PRIOR APPLICATION NUMBER: US 60/212,078
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 09/934,406
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: PCT/US01/26052
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: US 60/226,740
;  PRIOR FILING DATE: 2000-08-21
;  PRIOR APPLICATION NUMBER: US 09/861,801
;  PRIOR FILING DATE: 2001-05-21
;  PRIOR APPLICATION NUMBER: PCT/US01/16549
;  PRIOR FILING DATE: 2001-05-21
;  PRIOR APPLICATION NUMBER: US 60/205,508
;  PRIOR FILING DATE: 2000-05-19
;  PRIOR APPLICATION NUMBER: US 09/801,267
;  PRIOR FILING DATE: 2001-03-06
;  PRIOR APPLICATION NUMBER: PCT/US01/07138
;  PRIOR FILING DATE: 2001-03-05
;  PRIOR APPLICATION NUMBER: US 60/187,454
;  PRIOR FILING DATE: 2000-03-07
;  PRIOR APPLICATION NUMBER: US 09/829,671
;  PRIOR FILING DATE: 2001-04-10
;  PRIOR APPLICATION NUMBER: PCT/US01/40483
;  PRIOR FILING DATE: 2001-04-11
;  PRIOR APPLICATION NUMBER: US 60/197,508
;  PRIOR FILING DATE: 2000-04-18
;  PRIOR APPLICATION NUMBER: US 09/961,721
;  PRIOR FILING DATE: 2001-09-24
;  PRIOR APPLICATION NUMBER: PCT/US01/29904
;  PRIOR FILING DATE: 2001-09-24
;  PRIOR APPLICATION NUMBER: US 60/235,023
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US 10/045,367
;  PRIOR FILING DATE: 2001-11-07
;  PRIOR APPLICATION NUMBER: US 60/246,561
;  PRIOR FILING DATE: 2000-11-07
;  PRIOR APPLICATION NUMBER: US 09/801,275
;  PRIOR FILING DATE: 2001-03-06
;  PRIOR APPLICATION NUMBER: PCT/US01/07074
;  PRIOR FILING DATE: 2001-03-05
;  PRIOR APPLICATION NUMBER: US 60/187,420
;  PRIOR FILING DATE: 2000-03-07
;  NUMBER OF SEQ ID NOS: 63
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37
;   LENGTH: 3143
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (229)...(2874)
US-10-170-789-37

Alignment Scores:
Pred. No.:                1.29e-225       Length:          3143
Score:                    1836.50         Matches:         360
Percent Similarity:       67.54%          Conservative:    0
Best Local Similarity:    67.54%          Mismatches:      0
Query Match:              95.01%          Indels:          173
DB:                       12              Gaps:            1

US-10-070-464-7 (1-360) x US-10-170-789-37 (1-3143)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1017

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1077

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1137

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1197

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1198 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1257

Qy         98 ---------------------------------------------------------- 98

Db       1258 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1317

Qy         98 ---------------------------------------------------------- 98

Db       1318 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1377

Qy         98 ---------------------------------------------------------- 98

Db       1378 ACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1437

Qy         98 ---------------------------------------------------------- 98

Db       1438 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1497

Qy         98 ---------------------------------------------------------- 98

Db       1498 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1557

Qy         98 ---------------------------------------------------------- 98

Db       1558 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1617

Qy         98 ---------------------------------------------------------- 98

Db       1618 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1677

Qy         98 ---------------------------------------------------------- 98

Db       1678 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1737

Qy         99 ----------------------------------ValAspGluValArgArgLeuValTyr 107
                                                 |||||||||||||||||||||||||
Db       1738 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1797

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1798 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1857

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1858 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1917

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1918 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1977

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1978 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2037

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2038 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 2097

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2098 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 2157

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2158 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAAT 2217

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2218 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2277

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2278 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2337

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2338 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2397

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2398 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2457

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2458 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2517

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              |||||||||||||||||||||||||||||||||||||
Db       2518 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2556

RESULT 6
US-09-976-674-12
; Sequence 12, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
;   LENGTH: 4829
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-976-674-12

Alignment Scores:
Pred. No.:              9.6e-223      Length:          4829
Score:                  1816.50       Matches:         360
Percent Similarity:     67.29%        Conservative:    0
Best Local Similarity:  67.29%        Mismatches:      0
Query Match:            93.97%        Indels:          175
DB:                     10            Gaps:            1

US-10-070-464-7 (1-360) x US-09-976-674-12 (1-4829)

Qy           1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         943 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1002

Qy          21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1003 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1062

Qy          41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1063 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1122

Qy          61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        1123 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1182

Qy          81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1183 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1242

Qy          98 ---------------------------------------------------------- 98

Db        1243 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1302

Qy          98 ---------------------------------------------------------- 98

Db        1303 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1362

Qy          98 ---------------------------------------------------------- 98

Db        1363 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1422

Qy          98 ---------------------------------------------------------- 98

Db        1423 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1482

Qy          98 ---------------------------------------------------------- 98

Db        1483 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1542

Qy          98 ---------------------------------------------------------- 98

Db        1543 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1602

Qy          98 ---------------------------------------------------------- 98

Db        1603 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1662

Qy          98 ---------------------------------------------------------- 98

Db        1663 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1722

Qy          99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                 |||||||||||||||||||||||||
Db        1723 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1782

Qy         108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1783 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1842

Qy         128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1843 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1902

Qy         148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1903 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1962

Qy         168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1963 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2022

Qy         188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2023 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 2082

Qy         208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2083 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 2142

Qy         228 LysLysTyrProThrValLeuPheIleTyrGlyGly--ProGlnValGlnLeuValAsnA 247
               |||||||||||||||||||||||||||||||||||  |||||||||||||||||||||||
Db        2143 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCTCCTCAGGTGCAGTTGGTGAATA 2202

Qy         247 snArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValV 267
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2203 ATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGG 2262
```

```
Qy        267 alValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheL 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2263 TTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTA 2322

Qy        287 ysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaS 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2323 AATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTT 2382

Qy        307 erArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyT 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2383 CTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGAT 2442

Qy        327 yrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyA 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2443 ACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGG 2502

Qy        347 laProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||||||
Db       2503 CCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2543

RESULT 7
US-09-976-674-22
; Sequence 22, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
;   LENGTH: 4685
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-976-674-22

Alignment Scores:
Pred. No.:                8.42e-185       Length:         4685
Score:                    1522.00         Matches:        312
Percent Similarity:       58.54%          Conservative:   0
Best Local Similarity:    58.54%          Mismatches:     1
Query Match:              78.74%          Indels:         221
DB:                       10              Gaps:           2

US-10-070-464-7 (1-360) x US-09-976-674-22 (1-4685)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1002

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1062

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1122

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1182

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1242

Qy         98 ---------------------------------------------------------- 98
Db       1243 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1302

Qy         98 ---------------------------------------------------------- 98
Db       1303 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1362

Qy         98 ---------------------------------------------------------- 98
Db       1363 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1422

Qy         98 ---------------------------------------------------------- 98
Db       1423 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1482

Qy         98 ---------------------------------------------------------- 98
Db       1483 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1542

Qy         98 ---------------------------------------------------------- 98
Db       1543 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1602

Qy         98 ---------------------------------------------------------- 98
Db       1603 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1662

Qy         98 ---------------------------------------------------------- 98
Db       1663 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1722

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                |||||||||||||||||||||||||
Db       1723 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1782

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1842

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1902

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1962

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1963 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2022

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||
Db       2023 ATTTTGGATTCAGT---------------------------------------------- 2036

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2036 ---------------------------------------------------------- 2036

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
                                                  ||||||||||||||||||||||||
Db       2037 ----------------------------------CCTCAGGTGCAGTTGGTGAATAAT 2060

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2061 CGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTT 2120

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2121 GTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAA 2180
```

```
Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2181 TATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2240

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2241 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2300

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2301 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2360

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||
Db       2361 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2399

RESULT 8
US-09-976-674-20
; Sequence 20, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
;   LENGTH: 4676
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-976-674-20

Alignment Scores:
Pred. No.:              9.74e-185       Length:         4676
Score:                  1521.50         Matches:        309
Percent Similarity:     57.97%          Conservative:   0
Best Local Similarity:  57.97%          Mismatches:     1
Query Match:            78.71%          Indels:         224
DB:                     10              Gaps:           2

US-10-070-464-7 (1-360) x US-09-976-674-20 (1-4676)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1002

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1062

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1122

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1123 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1182

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1183 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1242

Qy         98 ------------------------------------------------------------ 98

Db       1243 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1302

Qy         98 ------------------------------------------------------------ 98

Db       1303 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1362

Qy         98 ------------------------------------------------------------ 98

Db       1363 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1422

Qy         98 ------------------------------------------------------------ 98

Db       1423 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1482

Qy         98 ------------------------------------------------------------ 98

Db       1483 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1542

Qy         98 ------------------------------------------------------------ 98

Db       1543 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1602

Qy         98 ------------------------------------------------------------ 98

Db       1603 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1662

Qy         98 ------------------------------------------------------------ 98

Db       1663 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1722

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107
                                                ||||||||||||||||||||||||
Db       1723 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1782

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1842

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1902

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1962

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1963 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2022

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2023 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 2082

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2083 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 2142

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247
              ||||||||||||||||||||||||||||||||||
Db       2143 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCG---------------------- 2180

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267

Db       2180 ------------------------------------------------------------ 2180

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287

Db       2180 ------------------------------------------------------------ 2180

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2181 ---------GGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATATCTAGCTTCT 2231

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327
```

```
Db       2232 CGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATAC 2291

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347

Db       2292 CTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATTGCTGGGGCC 2351

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360

Db       2352 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2390

RESULT 9
US-09-976-674-8
; Sequence 8, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
;   LENGTH: 4523
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-976-674-8

Alignment Scores:
Pred. No.:              4.69e-147       Length:         4523
Score:                  1229.00         Matches:        258
Percent Similarity:     48.41%          Conservative:   0
Best Local Similarity:  48.41%          Mismatches:     1
Query Match:            63.58%          Indels:         275
DB:                     10              Gaps:           2

US-10-070-464-7 (1-360) x US-09-976-674-8 (1-4523)

Qy           1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20

Db         943 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 1002

Qy          21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40

Db        1003 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 1062

Qy          41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60

Db        1063 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 1122

Qy          61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80

Db        1123 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACAGCAAATCCT 1182

Qy          81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98

Db        1183 AAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATCATAGATGTC 1242

Qy          98 ---------------------------------------------------------- 98

Db        1243 ATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAATATATTGCC 1302

Qy          98 ---------------------------------------------------------- 98

Db        1303 AGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGATCGCTCCCAG 1362

Qy          98 ---------------------------------------------------------- 98

Db        1363 ACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAAGATGATGTT 1422

Qy          98 ---------------------------------------------------------- 98

Db        1423 ATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTAATTATCTAT 1482

Qy          98 ---------------------------------------------------------- 98

Db        1483 GAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTTCCCCAAAGT 1542

Qy          98 ---------------------------------------------------------- 98

Db        1543 CACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTCCGTCATTTA 1602

Qy          98 ---------------------------------------------------------- 98

Db        1603 TACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGTGGGCTGCCT 1662

Qy          98 ---------------------------------------------------------- 98

Db        1663 GCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGG 1722

Qy          99 ---------------------------------ValAspGluValArgArgLeuValTyr 107

Db        1723 GAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATAT 1782

Qy         108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127

Db        1783 TTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAAT 1842

Qy         128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147

Db        1843 CCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAG 1902

Qy         148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167

Db        1903 CACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTT 1962

Qy         168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187

Db        1963 TACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACC 2022

Qy         188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207

Db        2023 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAA 2082

Qy         208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227

Db        2083 AGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGA 2142

Qy         228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247

Db        2143 AAGAAATATCCTACTGTGCTGTTCATATATGGTGGT---------------------- 2178

Qy         248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267

Db        2178 ---------------------------------------------------------- 2178

Qy         268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287

Db        2178 ---------------------------------------------------------- 2178

Qy         288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307

Db        2178 ---------------------------------------------------------- 2178

Qy         308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327

Db        2178 ---------------------------------------------------------- 2178

Qy         328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347
```

```
Db       2179 ------------------------------------CGGTT-GCTATTGCTGGGGCC 2198

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||||
Db       2199 CCAGTCACTCTGTGGATCTTCTATGATACAGGATACACG 2237

RESULT 10
US-09-976-674-4
; Sequence 4, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 2617
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-976-674-4

Alignment Scores:
Pred. No.:              2.46e-146     Length:          2617
Score:                  1220.50       Matches:         236
Percent Similarity:     54.31%        Conservative:    54
Best Local Similarity:  44.19%        Mismatches:      69
Query Match:            63.14%        Indels:          175
DB:                     10            Gaps:            2

US-10-070-464-7 (1-360) x US-09-976-674-4 (1-2617)

Qy          2 GluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArgTyr 21
              :::|||   :::||||||||||||||||||||||||:::||||||||||||||||||:::
Db        710 GATGACCCCAAGTCTGCGGGTGTGGCCACCTTCGTCATACAGGAAGAGTTCGACCGCTTC 769

Qy         22 SerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly---LysIleLeu 40
              :::||||||||||||||||||   |||          |||   |||   |||   |||
Db        770 ACTGGGTACTGGTGGTGCCCCACAGCCTCCTGGGAAGGTTCAGAGGGCCTCAAGACGCTG 829

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||   ||||||||||||||||||:::|||||||||   ||||||
Db        830 CGAATCCTGTATGAGGAAGTCGATGAGTCCGAGGTGGAGGTCATTCACGTCCCCTCTCCT 889

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
                 ||||||   |||:::   ||||||:::||||||||:::||||||:::   ||||||
Db        890 GCGCTAGAAGAAAGGAAGACGGACTCGTATCGGTACCCCAGGACAGGCAGCAAGAATCCC 949

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              |||:::      |||::::::|||         |||::::::|||:::|||:::
Db        950 AAGATTGCCTTGAAACTGGCTGAGTTCCAGACTGACAGCCAGGGCAAGATCGTCTCGACC 1009

Qy         98 ------------------------------------------------------------ 98

Db       1010 CAGGAGAAGGAGCTGGTGCAGCCCTTCAGCTCGCTGTTCCCGAAGGTGGAGTACATCGCC 1069

Qy         98 ------------------------------------------------------------ 98

Db       1070 AGGGCCGGGTGGACCCGGGATGGCAAATACGCCTGGGCCATGTTCCTGGACCGGCCCCAG 1129

Qy         98 ------------------------------------------------------------ 98

Db       1130 CAGTGGCTCCAGCTCGTCCTCCTCCCCCCGGCCCTGTTCATCCCGAGCACAGAGAATGAG 1189

Qy         98 ------------------------------------------------------------ 98

Db       1190 GAGCAGCGGCTAGCCTCTGCCAGAGCTGTCCCCAGGAATGTCCAGCCGTATGTGGTGTAC 1249

Qy         98 ------------------------------------------------------------ 98

Db       1250 GAGGAGGTCACCAACGTCTGGATCAATGTTCATGACATCTTCTATCCCTTCCCCCAATCA 1309

Qy         98 ------------------------------------------------------------ 98

Db       1310 GAGGGAGAGGACGAGCTCTGCTTTCTCCGCGCCAATGAATGCAAGACCGGCTTCTGCCAT 1369

Qy         98 ------------------------------------------------------------ 98

Db       1370 TTGTACAAAGTCACCGCCGTTTTAAAATCCCAGGGCTACGATTGGAGTGAGCCCTTCAGC 1429

Qy         98 ------------------------------------------------------------ 98

Db       1430 CCCGGGGAAGATGAATTTAAGTGCCCCATTAAGGAAGAGATTGCTCTGACCAGCGGTGAA 1489

Qy         99 ----------------------------------ValAspGluValArgArgLeuVal 106
                                                 |||:::|||      :::||||||
Db       1490 TGGGAGGTTTTGGCGAGGCACGGCTCCAAGATCTGGGTCAATGAGGAGACCAAGCTGGTG 1549

Qy        107 TyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrVal 126
              ||||||:::|||||||||||:::||||||||||||||||||||||||||||||||||||
Db       1550 TACTTCCAGGGCACCAAGGACACGCCGCTGGAGCACCACCTCTACGTGGTCAGCTATGAG 1609

Qy        127 AsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSer 146
                    ||||||:::   |||||||||      |||:::||||||||||||   :::|||
Db       1610 GCGGCCGGCGAGATCGTACGCCTCACCACGCCCGGCTTCTCCCATAGCTGCTCCATGAGC 1669

Qy        147 GlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSer 166
              |||:::   |||   |||:::|||   ||||||:::          |||   ||||||
Db       1670 CAGAACTTCGACATGTTCGTCAGCCACTACAGCAGCGTGAGCACGCCGCCCTGCGTGCAC 1729

Qy        167 LeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAla 186
              :::||||||||||||   |||:::|||||||||      |||          |||||||||
Db       1730 GTCTACAAGCTGAGCGGCCCCGACGACGACCCCCTGCACAAGCAGCCCCGCTTCTGGGCT 1789

Qy        187 ThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPhe 206
              :::::::::::::::|||          |||||||||   ||||||||||||||   |||
Db       1790 AGCATGATGGAGGCAGCCAGCTGCCCCCCGGATTATGTTCCTCCAGAGATCTTCCATTTC 1849

Qy        207 GluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnPro 226
              :::   :::          ||||||||||||:::||||||||||||   |||||||||
Db       1850 CACACGCGCTCGGATGTGCGGCTCTACGGCATGATCTACAAGCCCCACGCCTTGCAGCCA 1909

Qy        227 GlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsn 246
              |||||||||:::||||||||||||||:::||||||||||||||||||||||||||||||
Db       1910 GGGAAGAAGCACCCCACCGTCCTCTTTGTATATGGAGGCCCCCAGGTGCAGCTGGTGAAT 1969

Qy        247 AsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrVal 266
              |||   |||||||||:::||||||   ||||||||||||||||||||||||||||||
Db       1970 AACTCCTTCAAAGGCATCAAGTACTTGCGGCTCAACACACTGGCCTCCCTGGGCTACGCC 2029

Qy        267 ValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPhe 286
              ||||||||||||||   ||||||||||   |||||||||:::||||||||||||||||
Db       2030 GTGGTTGTGATTGACGGCAGGGGCTCCTGTCAGCGAGGGCTTCGGTTCGAAGGGGCCCTG 2089

Qy        287 LysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAla 306
              |||   :::|||||||:::||||||:::||||||||||||||||||||||:::::|||
Db       2090 AAAAACCAAATGGGCCAGGTGGAGATCGAGGACCAGGTGGAGGGCCTGCAGTTCGTGGCC 2149

Qy        307 SerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGly 326
                 :::|||   ||||||||||||   ||||||   ||||||||||||||||||||||||
Db       2150 GAGAAGTATGGCTTCATCGACCTGAGCCGAGTTGCCATCCATGGCTGGTCCTACGGGGGC 2209

Qy        327 TyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGly 346
              :::||||||||||||   |||:::   :::      :::|||:::||||||||||||||
Db       2210 TTCCTCTCGCTCATGGGGCTAATCCACAAGCCCCAGGTGTTCAAGGTGGCCATCGCGGGT 2269

Qy        347 AlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||:::|||:::   ||||||||||||||||||
Db       2270 GCCCCGGTCACCGTCTGGATGGCCTACGACACAGGGTACACT 2311
```

```
RESULT 11
US-09-976-674-28
; Sequence 28, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
;   LENGTH: 4219
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-976-674-28

Alignment Scores:
Pred. No.:              5.24e-146     Length:          4219
Score:                  1220.50       Matches:         236
Percent Similarity:     54.31%        Conservative:    54
Best Local Similarity:  44.19%        Mismatches:      69
Query Match:            63.14%        Indels:          175
DB:                     10            Gaps:            2

US-10-070-464-7 (1-360) x US-09-976-674-28 (1-4219)

Qy          2 GluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArgTyr 21
              :::|||   :::||||||||||||||||||||||||:::|||||||||||||||||||:::
Db       1066 GATGACCCCAAGTCTGCGGGTGTGGCCACCTTCGTCATACAGGAAGAGTTCGACCGCTTC 1125

Qy         22 SerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly---LysIleLeu 40
              :::||||||||||||||||||  |||        |||  |||  |||   |||
Db       1126 ACTGGGTACTGGTGGTGCCCCACAGCCTCCTGGGAAGGTTCAGAGGGCCTCAAGACGCTG 1185

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||  ||||||||||||||||||:::|||||||||  ||||||
Db       1186 CGAATCCTGTATGAGGAAGTCGATGAGTCCGAGGTGGAGGTCATTCACGTCCCCTCTCCT 1245

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
                ||||||   |||:::   |||||||:::|||||||:::|||||||:::   ||||||
Db       1246 GCGCTAGAAGAAAGGAAGACGGACTCGTATCGGTACCCCAGGACAGGCAGCAAGAATCCC 1305

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              |||:::      |||:::::::|||        |||:::::::|||:::|||:::
Db       1306 AAGATTGCCTTGAAACTGGCTGAGTTCCAGACTGACAGCCAGGGCAAGATCGTCTCGACC 1365

Qy         98 ------------------------------------------------------------ 98

Db       1366 CAGGAGAAGGAGCTGGTGCAGCCCTTCAGCTCGCTGTTCCCGAAGGTGGAGTACATCGCC 1425

Qy         98 ------------------------------------------------------------ 98

Db       1426 AGGGCCGGGTGGACCCGGGATGGCAAATACGCCTGGGCCATGTTCCTGGACCGGCCCCAG 1485

Qy         98 ------------------------------------------------------------ 98

Db       1486 CAGTGGCTCCAGCTCGTCCTCCTCCCCCCGGCCCTGTTCATCCCGAGCACAGAGAATGAG 1545

Qy         98 ------------------------------------------------------------ 98

Db       1546 GAGCAGCGGCTAGCCTCTGCCAGAGCTGTCCCCAGGAATGTCCAGCCGTATGTGGTGTAC 1605

Qy         98 ------------------------------------------------------------ 98

Db       1606 GAGGAGGTCACCAACGTCTGGATCAATGTTCATGACATCTTCTATCCCTTCCCCCAATCA 1665

Qy         98 ------------------------------------------------------------ 98

Db       1666 GAGGGAGAGGACGAGCTCTGCTTTCTCCGCGCCAATGAATGCAAGACCGGCTTCTGCCAT 1725

Qy         98 ------------------------------------------------------------ 98

Db       1726 TTGTACAAAGTCACCGCCGTTTTAAAATCCCAGGGCTACGATTGGAGTGAGCCCTTCAGC 1785

Qy         98 ------------------------------------------------------------ 98

Db       1786 CCCGGGGAAGATGAATTTAAGTGCCCCATTAAGGAAGAGATTGCTCTGACCAGCGGTGAA 1845

Qy         99 ----------------------------------ValAspGluValArgArgLeuVal 106
                                                 |||:::|||      :::||||||
Db       1846 TGGGAGGTTTTGGCGAGGCACGGCTCCAAGATCTGGGTCAATGAGGAGACCAAGCTGGTG 1905

Qy        107 TyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrVal 126
              ||||||:::|||||||||||||:::||||||||||||||||||||||||||||||||||
Db       1906 TACTTCCAGGGCACCAAGGACACGCCGCTGGAGCACCACCTCTACGTGGTCAGCTATGAG 1965

Qy        127 AsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSer 146
                 ||||||:::   |||||||||     |||:::||||||||||||   :::|||
Db       1966 GCGGCCGGCGAGATCGTACGCCTCACCACGCCCGGCTTCTCCCATAGCTGCTCCATGAGC 2025

Qy        147 GlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSer 166
              |||:::  |||  |||:::|||  ||||||:::     |||  ||||||
Db       2026 CAGAACTTCGACATGTTCGTCAGCCACTACAGCAGCGTGAGCACGCCGCCCTGCGTGCAC 2085

Qy        167 LeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAla 186
              :::||||||||||||  |||:::|||||||||     |||         |||||||||
Db       2086 GTCTACAAGCTGAGCGGCCCCGACGACGACCCCCTGCACAAGCAGCCCCGCTTCTGGGCT 2145

Qy        187 ThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPhe 206
              ::::::::::::::|||         |||||||||  ||||||||||||||   |||
Db       2146 AGCATGATGGAGGCAGCCAGCTGCCCCCCGGATTATGTTCCTCCAGAGATCTTCCATTTC 2205

Qy        207 GluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnPro 226
              :::  :::          ||||||||||||:::||||||||||||  |||||||||
Db       2206 CACACGCGCTCGGATGTGCGGCTCTACGGCATGATCTACAAGCCCCACGCCTTGCAGCCA 2265

Qy        227 GlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsn 246
              |||||||||:::|||||||||||||:::|||||||||||||||||||||||||||||||
Db       2266 GGGAAGAAGCACCCCACCGTCCTCTTTGTATATGGAGGCCCCCAGGTGCAGCTGGTGAAT 2325

Qy        247 AsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrVal 266
              |||   |||||||||:::|||||   ||||||||||||||||||||||||||||||
Db       2326 AACTCCTTCAAAGGCATCAAGTACTTGCGGCTCAACACACTGGCCTCCCTGGGCTACGCC 2385

Qy        267 ValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPhe 286
              ||||||||||||||   ||||||||||||   ||||||||:::||||||||||||
Db       2386 GTGGTTGTGATTGACGGCAGGGGCTCCTGTCAGCGAGGGCTTCGGTTCGAAGGGGCCCTG 2445

Qy        287 LysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAla 306
              |||   :::||||||||||:::||||||:::||||||||||||||||||:::::|||
Db       2446 AAAAACCAAATGGGCCAGGTGGAGATCGAGGACCAGGTGGAGGGCCTGCAGTTCGTGGCC 2505

Qy        307 SerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGly 326
                 :::|||   ||||||||||||   ||||||   ||||||||||||||||||||||||
Db       2506 GAGAAGTATGGCTTCATCGACCTGAGCCGAGTTGCCATCCATGGCTGGTCCTACGGGGGC 2565

Qy        327 TyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGly 346
              :::||||||||||||   |||:::   :::        :::|||:::||||||||||||
Db       2566 TTCCTCTCGCTCATGGGGCTAATCCACAAGCCCCAGGTGTTCAAGGTGGCCATCGCGGGT 2625

Qy        347 AlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||:::|||:::   ||||||||||||||||
Db       2626 GCCCCGGTCACCGTCTGGATGGCCTACGACACAGGGTACACT 2667

RESULT 12
US-09-976-674-24
; Sequence 24, Application US/09976674
; Patent No. US20020115843A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
;   LENGTH: 4302
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-976-674-24

Alignment Scores:
Pred. No.:                5.4e-146      Length:          4302
Score:                    1220.50       Matches:         236
Percent Similarity:       54.31%        Conservative:    54
Best Local Similarity:    44.19%        Mismatches:      69
Query Match:              63.14%        Indels:          175
DB:                       10            Gaps:            2

US-10-070-464-7 (1-360) x US-09-976-674-24 (1-4302)

Qy          2 GluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArgTyr 21
              :::|||   :::|||||||||||||||||||||||||:::|||||||||||||||||:::
Db       1066 GATGACCCCAAGTCTGCGGGTGTGGCCACCTTCGTCATACAGGAAGAGTTCGACCGCTTC 1125

Qy         22 SerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly---LysIleLeu 40
              :::|||||||||||||||||   |||          |||   |||   |||   |||
Db       1126 ACTGGGTACTGGTGGTGCCCCACAGCCTCCTGGGAAGGTTCAGAGGGCCTCAAGACGCTG 1185

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              |||||||||||||||||   |||||||||||||||||:::|||||||||   ||||||
Db       1186 CGAATCCTGTATGAGGAAGTCGATGAGTCCGAGGTGGAGGTCATTCACGTCCCCTCTCCT 1245

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
                 ||||||   |||:::   ||||||:::|||||||||:::|||||||:::   ||||||
Db       1246 GCGCTAGAAGAAAGGAAGACGGACTCGTATCGGTACCCCAGGACAGGCAGCAAGAATCCC 1305

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              |||:::      |||:::::::|||         |||:::::::|||:::|||:::
Db       1306 AAGATTGCCTTGAAACTGGCTGAGTTCCAGACTGACAGCCAGGGCAAGATCGTCTCGACC 1365

Qy         98 ---------------------------------------------------------- 98

Db       1366 CAGGAGAAGGAGCTGGTGCAGCCCTTCAGCTCGCTGTTCCCGAAGGTGGAGTACATCGCC 1425

Qy         98 ---------------------------------------------------------- 98

Db       1426 AGGGCCGGGTGGACCCGGGATGGCAAATACGCCTGGGCCATGTTCCTGGACCGGCCCCAG 1485

Qy         98 ---------------------------------------------------------- 98

Db       1486 CAGTGGCTCCAGCTCGTCCTCCTCCCCCCGGCCCTGTTCATCCCGAGCACAGAGAATGAG 1545

Qy         98 ---------------------------------------------------------- 98

Db       1546 GAGCAGCGGCTAGCCTCTGCCAGAGCTGTCCCCAGGAATGTCCAGCCGTATGTGGTGTAC 1605

Qy         98 ---------------------------------------------------------- 98

Db       1606 GAGGAGGTCACCAACGTCTGGATCAATGTTCATGACATCTTCTATCCCTTCCCCCAATCA 1665

Qy         98 ---------------------------------------------------------- 98

Db       1666 GAGGGAGAGGACGAGCTCTGCTTTCTCCGCGCCAATGAATGCAAGACCGGCTTCTGCCAT 1725

Qy         98 ---------------------------------------------------------- 98

Db       1726 TTGTACAAAGTCACCGCCGTTTTAAAATCCCAGGGCTACGATTGGAGTGAGCCCTTCAGC 1785

Qy         98 ---------------------------------------------------------- 98

Db       1786 CCCGGGGAAGATGAATTTAAGTGCCCCATTAAGGAAGAGATTGCTCTGACCAGCGGTGAA 1845

Qy         99 ----------------------------------ValAspGluValArgArgLeuVal 106
                                                |||:::|||      :::|||||
Db       1846 TGGGAGGTTTTGGCGAGGCACGGCTCCAAGATCTGGGTCAATGAGGAGACCAAGCTGGTG 1905

Qy        107 TyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrVal 126
              ||||||:::|||||||||||:::|||||||||||||||||||||||||||||||||||
Db       1906 TACTTCCAGGGCACCAAGGACACGCCGCTGGAGCACCACCTCTACGTGGTCAGCTATGAG 1965

Qy        127 AsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSer 146
                    |||||||:::   |||||||||      |||:::||||||||||||   :::|||
Db       1966 GCGGCCGGCGAGATCGTACGCCTCACCACGCCCGGCTTCTCCCATAGCTGCTCCATGAGC 2025

Qy        147 GlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSer 166
              |||:::   |||   |||:::|||   ||||||:::         |||   ||||||
Db       2026 CAGAACTTCGACATGTTCGTCAGCCACTACAGCAGCGTGAGCACGCCGCCCTGCGTGCAC 2085

Qy        167 LeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAla 186
              :::|||||||||||||   |||:::||||||||||      |||           |||||||||
Db       2086 GTCTACAAGCTGAGCGGCCCCGACGACGACCCCCTGCACAAGCAGCCCCGCTTCTGGGCT 2145

Qy        187 ThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPhe 206
              :::::::::::::::|||          |||||||||   ||||||||||||||   |||
Db       2146 AGCATGATGGAGGCAGCCAGCTGCCCCCCGGATTATGTTCCTCCAGAGATCTTCCATTTC 2205

Qy        207 GluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnPro 226
                 :::   :::          ||||||||||||:::|||||||||||   |||||||||
Db       2206 CACACGCGCTCGGATGTGCGGCTCTACGGCATGATCTACAAGCCCCACGCCTTGCAGCCA 2265

Qy        227 GlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsn 246
              ||||||||||:::|||||||||||||||:::|||||||||||||||||||||||||||||
Db       2266 GGGAAGAAGCACCCCACCGTCCTCTTTGTATATGGAGGCCCCCAGGTGCAGCTGGTGAAT 2325

Qy        247 AsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrVal 266
              |||   ||||||||||:::|||||   ||||||||||||||||||||||||||||||||
Db       2326 AACTCCTTCAAAGGCATCAAGTACTTGCGGCTCAACACACTGGCCTCCCTGGGCTACGCC 2385

Qy        267 ValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPhe 286
              |||||||||||||||   ||||||||||||   |||||||||:::|||||||||||||
Db       2386 GTGGTTGTGATTGACGGCAGGGGCTCCTGTCAGCGAGGGCTTCGGTTCGAAGGGGCCCTG 2445

Qy        287 LysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAla 306
              |||   :::|||||||||:::|||||||:::||||||||||||||||||:::::|||
Db       2446 AAAAACCAAATGGGCCAGGTGGAGATCGAGGACCAGGTGGAGGGCCTGCAGTTCGTGGCC 2505

Qy        307 SerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGly 326
                 :::|||   ||||||||||||   ||||||   ||||||||||||||||||||||
Db       2506 GAGAAGTATGGCTTCATCGACCTGAGCCGAGTTGCCATCCATGGCTGGTCCTACGGGGGC 2565

Qy        327 TyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGly 346
              :::|||||||||||||   |||:::   :::      :::|||:::|||||||||||||||
Db       2566 TTCCTCTCGCTCATGGGGCTAATCCACAAGCCCCAGGTGTTCAAGGTGGCCATCGCGGGT 2625

Qy        347 AlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              |||||||||||:::|||:::   |||||||||||||||||
Db       2626 GCCCCGGTCACCGTCTGGATGGCCTACGACACAGGGTACACT 2667

RESULT 13
US-09-976-674-36
; Sequence 36, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
```

```
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36
;   LENGTH: 4180
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-976-674-36

Alignment Scores:
Pred. No.:                1.69e-143      Length:          4180
Score:                    1201.00        Matches:         234
Percent Similarity:       55.09%         Conservative:    53
Best Local Similarity:    44.91%         Mismatches:      72
Query Match:              62.13%         Indels:          162
DB:                       10             Gaps:            3

US-10-070-464-7 (1-360) x US-09-976-674-36 (1-4180)

Qy          2 GluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArgTyr 21
              :::|||   :::|||||||||||||||||||||||||::|||||||||||||||||||:::
Db       1066 GATGACCCCAAGTCTGCGGGTGTGGCCACCTTCGTCATACAGGAAGAGTTCGACCGCTTC 1125

Qy         22 SerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly---LysIleLeu 40
              :::|||||||||||||||||   |||           |||   |||   |||   |||
Db       1126 ACTGGGTACTGGTGGTGCCCCACAGCCTCCTGGGAAGGTTCAGAGGGCCTCAAGACGCTG 1185

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||   ||||||||||||||||:::|||||||||   ||||||
Db       1186 CGAATCCTGTATGAGGAAGTCGATGAGTCCGAGGTGGAGGTCATTCACGTCCCCTCTCCT 1245

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
                ||||||   |||:::   ||||||:::|||||||||:::|||||||:::   ||||||
Db       1246 GCGCTAGAAGAAAGGAAGACGGACTCGTATCGGTACCCCAGGACAGGCAGCAAGAATCCC 1305

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIleValAsp 100
              |||:::      |||:::::::|||         |||:::::::|||:::|||:::
Db       1306 AAGATTGCCTTGAAACTGGCTGAGTTCCAGACTGACAGCCAGGGCAAGATCGTCTCGACC 1365

Qy        101 GluValArgArgLeu---------------------------------------------- 105
              :::   :::   |||
Db       1366 CAGGAGAAGGAGCTGGTGCAGCCCTTCAGCTCGCTGTTCCCGAAGGTGGAGTACATCGCC 1425

Qy        105 ---------------------------------------------------------- 105

Db       1426 AGGGCCGGGTGGACCCGGGATGGCAAATACGCCTGGGCCATGTTCCTGGACCGGCCCCAG 1485

Qy        105 ---------------------------------------------------------- 105

Db       1486 CAGTGGCTCCAGCTCGTCCTCCTCCCCCCGGCCCTGTTCATCCCGAGCACAGAGAATGAG 1545

Qy        106 ----------------------------------------------------------ValTyr 107
                                                                      ||||||
Db       1546 GAGCAGCGGCTAGCCTCTGCCAGAGCTGTCCCCAGGAATGTCCAGCCGTATGTGGTGTAC 1605

Qy        108 PheGlu---------------------------------------------------- 109
              |||
Db       1606 GAGGAGGTCACCAACGTCTGGATCAATGTTCATGACATCTTCTATCCCTTCCCCCAATCA 1665

Qy        109 ---------------------------------------------------------- 109

Db       1666 GAGGGAGAGGACGAGCTCTGCTTTCTCCGCGCCAATGAATGCAAGACCGGCTTCTGCCAT 1725

Qy        109 ---------------------------------------------------------- 109

Db       1726 TTGTACAAAGTCACCGCCGTTTTAAAATCCCAGGGCTACGATTGGAGTGAGCCCTTCAGC 1785

Qy        109 ---------------------------------------------------------- 109

Db       1786 CCCGGGGAAGATGAATTTAAGTGCCCCATTAAGGAAGAGATTGCTCTGACCAGCGGTGAA 1845

Qy        110 ----------------------------GlyThrLysAspSerProLeuGluHisHis 119
                                          ||||||||||||:::|||||||||||||||
Db       1846 TGGGAGGTTTTGGCGAGGCACGGCTCCAAGGGCACCAAGGACACGCCGCTGGAGCACCAC 1905

Qy        120 LeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyr 139
              ||||||||||||||||||          ||||||:::   |||||||||      |||:::
Db       1906 CTCTACGTGGTCAGCTATGAGGCGGCCGGCGAGATCGTACGCCTCACCACGCCCGGCTTC 1965

Qy        140 SerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGln 159
              ||||||||||||   :::||||||:::   |||   |||:::|||   ||||||:::
Db       1966 TCCCATAGCTGCTCCATGAGCCAGAACTTCGACATGTTCGTCAGCCACTACAGCAGCGTG 2025

Qy        160 LysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCys 179
                 |||   ||||||   :::|||||||||||   |||:::|||||||||
Db       2026 AGCACGCCGCCCTGCGTGCACGTCTACAAGCTGAGCGGCCCCGACGACGACCCCCTGCAC 2085

Qy        180 LysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThr 199
              |||         |||||||||:::::::::::::::|||         |||||||||
Db       2086 AAGCAGCCCCGCTTCTGGGCTAGCATGATGGAGGCAGCCAGCTGCCCCCCGGATTATGTT 2145

Qy        200 ProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyr 219
              ||||||||||||||||   |||   :::   :::          ||||||||||||:::|||
Db       2146 CCTCCAGAGATCTTCCATTTCCACACGCGCTCGGATGTGCGGCTCTACGGCATGATCTAC 2205

Qy        220 LysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly 239
              |||||||||   ||||||||||||||||||:::|||||||||||||||:::|||||||||
Db       2206 AAGCCCCACGCCTTGCAGCCAGGGAAGAAGCACCCCACCGTCCTCTTTGTATATGGAGGC 2265

Qy        240 ProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThr 259
              ||||||||||||||||||||||||   |||||||||:::|||||||   ||||||||||||
Db       2266 CCCCAGGTGCAGCTGGTGAATAACTCCTTCAAAGGCATCAAGTACTTGCGGCTCAACACA 2325

Qy        260 LeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCysHisArgGly 279
              ||||||||||||||||||   ||||||||||||||   ||||||||||||   ||||||
Db       2326 CTGGCCTCCCTGGGCTACGCCGTGGTTGTGATTGACGGCAGGGGCTCCTGTCAGCGAGGG 2385

Qy        280 LeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnVal 299
              |||:::|||||||||||||   |||   :::|||||||||:::|||||||:::|||||||||
Db       2386 CTTCGGTTCGAAGGGGCCCTGAAAAACCAAATGGGCCAGGTGGAGATCGAGGACCAGGTG 2445

Qy        300 GluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIle 319
              ||||||||||||::::::|||   :::|||   ||||||||||||   ||||||   |||
Db       2446 GAGGGCCTGCAGTTCGTGGCCGAGAAGTATGGCTTCATCGACCTGAGCCGAGTTGCCATC 2505

Qy        320 HisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIle 339
              ||||||||||||||||||:::||||||||||||   |||:::   :::       :::
Db       2506 CATGGCTGGTCCTACGGGGGCTTCCTCTCGCTCATGGGGCTAATCCACAAGCCCCAGGTG 2565

Qy        340 PheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyr 359
              |||:::|||||||||||||||||||||||||:::|||:::   ||||||||||||||||
Db       2566 TTCAAGGTGGCCATCGCGGGTGCCCCGGTCACCGTCTGGATGGCCTACGACACAGGGTAC 2625

Qy        360 Thr 360
              |||
Db       2626 ACT 2628

RESULT 14
US-09-976-674-34
; Sequence 34, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
```

```
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
;   LENGTH: 4263
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-976-674-34

Alignment Scores:
Pred. No.:                1.74e-143      Length:          4263
Score:                    1201.00        Matches:         234
Percent Similarity:       55.09%         Conservative:    53
Best Local Similarity:    44.91%         Mismatches:      72
Query Match:              62.13%         Indels:          162
DB:                       10             Gaps:            3

US-10-070-464-7 (1-360) x US-09-976-674-34 (1-4263)

Qy          2 GluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArgTyr 21
              :::|||   :::|||||||||||||||||||||||||:::|||||||||||||||||:::
Db       1066 GATGACCCCAAGTCTGCGGGTGTGGCCACCTTCGTCATACAGGAAGAGTTCGACCGCTTC 1125

Qy         22 SerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly---LysIleLeu 40
              :::|||||||||||||||||||   |||             |||   |||   |||   |||
Db       1126 ACTGGGTACTGGTGGTGCCCCACAGCCTCCTGGGAAGGTTCAGAGGGCCTCAAGACGCTG 1185

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||   |||||||||||||||||:::|||||||||   ||||||
Db       1186 CGAATCCTGTATGAGGAAGTCGATGAGTCCGAGGTGGAGGTCATTCACGTCCCCTCTCCT 1245

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
                ||||||   |||:::   ||||||:::|||||||||||:::||||||:::   ||||||
Db       1246 GCGCTAGAAGAAAGGAAGACGGACTCGTATCGGTACCCCAGGACAGGCAGCAAGAATCCC 1305

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIleValAsp 100
              |||:::      |||:::::|||         |||:::::|||:::|||:::
Db       1306 AAGATTGCCTTGAAACTGGCTGAGTTCCAGACTGACAGCCAGGGCAAGATCGTCTCGACC 1365

Qy        101 GluValArgArgLeu---------------------------------------------- 105
              :::   :::   |||
Db       1366 CAGGAGAAGGAGCTGGTGCAGCCCTTCAGCTCGCTGTTCCCGAAGGTGGAGTACATCGCC 1425

Qy        105 ---------------------------------------------------------- 105

Db       1426 AGGGCCGGGTGGACCCGGGATGGCAAATACGCCTGGGCCATGTTCCTGGACCGGCCCCAG 1485

Qy        105 ---------------------------------------------------------- 105

Db       1486 CAGTGGCTCCAGCTCGTCCTCCTCCCCCCGGCCCTGTTCATCCCGAGCACAGAGAATGAG 1545

Qy        106 ----------------------------------------------------------ValTyr 107
                                                                       ||||||
Db       1546 GAGCAGCGGCTAGCCTCTGCCAGAGCTGTCCCCAGGAATGTCCAGCCGTATGTGGTGTAC 1605

Qy        108 PheGlu---------------------------------------------------- 109
              |||
Db       1606 GAGGAGGTCACCAACGTCTGGATCAATGTTCATGACATCTTCTATCCCTTCCCCCAATCA 1665

Qy        109 ---------------------------------------------------------- 109

Db       1666 GAGGGAGAGGACGAGCTCTGCTTTCTCCGCGCCAATGAATGCAAGACCGGCTTCTGCCAT 1725

Qy        109 ---------------------------------------------------------- 109

Db       1726 TTGTACAAAGTCACCGCCGTTTTAAAATCCCAGGGCTACGATTGGAGTGAGCCCTTCAGC 1785

Qy        109 ---------------------------------------------------------- 109

Db       1786 CCCGGGGAAGATGAATTTAAGTGCCCCATTAAGGAAGAGATTGCTCTGACCAGCGGTGAA 1845

Qy        110 ----------------------------GlyThrLysAspSerProLeuGluHisHis 119
                                          ||||||||||||:::||||||||||||||||

Db       1846 TGGGAGGTTTTGGCGAGGCACGGCTCCAAGGGCACCAAGGACACGCCGCTGGAGCACCAC 1905

Qy        120 LeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyr 139
              ||||||||||||||||||          ||||||:::   |||||||||      |||:::
Db       1906 CTCTACGTGGTCAGCTATGAGGCGGCCGGCGAGATCGTACGCCTCACCACGCCCGGCTTC 1965

Qy        140 SerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGln 159
              ||||||||||||   :::||||||:::   |||   |||:::|||   ||||||:::
Db       1966 TCCCATAGCTGCTCCATGAGCCAGAACTTCGACATGTTCGTCAGCCACTACAGCAGCGTG 2025

Qy        160 LysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCys 179
              |||   ||||||   :::||||||||||||   |||:::|||||||||
Db       2026 AGCACGCCGCCCTGCGTGCACGTCTACAAGCTGAGCGGCCCCGACGACGACCCCCTGCAC 2085

Qy        180 LysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThr 199
              |||          |||||||||:::::::::::::::|||          |||||||||
Db       2086 AAGCAGCCCCGCTTCTGGGCTAGCATGATGGAGGCAGCCAGCTGCCCCCCGGATTATGTT 2145

Qy        200 ProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyr 219
              ||||||||||||||   |||   :::   :::        ||||||||||||:::|||
Db       2146 CCTCCAGAGATCTTCCATTTCCACACGCGCTCGGATGTGCGGCTCTACGGCATGATCTAC 2205

Qy        220 LysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly 239
              |||||||||   |||||||||||||||||:::|||||||||||||||:::|||||||||
Db       2206 AAGCCCCACGCCTTGCAGCCAGGGAAGAAGCACCCCACCGTCCTCTTTGTATATGGAGGC 2265

Qy        240 ProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThr 259
              ||||||||||||||||||||||||   ||||||||:::||||||   ||||||||||||
Db       2266 CCCCAGGTGCAGCTGGTGAATAACTCCTTCAAAGGCATCAAGTACTTGCGGCTCAACACA 2325

Qy        260 LeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCysHisArgGly 279
              ||||||||||||||||||   ||||||||||||||   ||||||||||||   ||||||
Db       2326 CTGGCCTCCCTGGGCTACGCCGTGGTTGTGATTGACGGCAGGGGCTCCTGTCAGCGAGGG 2385

Qy        280 LeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnVal 299
              |||:::|||||||||||   |||   :::|||||||||:::||||||:::||||||||||
Db       2386 CTTCGGTTCGAAGGGGCCCTGAAAAACCAAATGGGCCAGGTGGAGATCGAGGACCAGGTG 2445

Qy        300 GluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIle 319
              ||||||||||||:::::|||   :::|||   ||||||||||||   ||||||   |||
Db       2446 GAGGGCCTGCAGTTCGTGGCCGAGAAGTATGGCTTCATCGACCTGAGCCGAGTTGCCATC 2505

Qy        320 HisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIle 339
              ||||||||||||||||||||:::||||||||||||   |||:::   :::      :::
Db       2506 CATGGCTGGTCCTACGGGGGCTTCCTCTCGCTCATGGGGCTAATCCACAAGCCCCAGGTG 2565

Qy        340 PheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyr 359
              |||:::|||||||||||||||||||||||||||||||:::|||:::   ||||||||||||
Db       2566 TTCAAGGTGGCCATCGCGGGTGCCCCGGTCACCGTCTGGATGGCCTACGACACAGGGTAC 2625

Qy        360 Thr 360
              |||
Db       2626 ACT 2628

RESULT 15
US-10-098-841-83
; Sequence 83, Application US/10098841
; Publication No. US20020197679A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Liu, Chenghua
;  APPLICANT:  Asundi, Vinod
;  APPLICANT:  Xu, Chongjun
;  APPLICANT:  Zhou, Ping
;  APPLICANT:  Ma, Yunqing
;  APPLICANT:  Wang, Jian-Rui
;  APPLICANT:  Zhao, Qing A.
;  APPLICANT:  Ren, Feiyan
;  APPLICANT:  Chen, Rui-hong
;  APPLICANT:  Wang, Dunrui
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Wehrman, Tom
;  APPLICANT:  Zhang, Jie
```

```
;  APPLICANT:  Qian, Xiaohong B.
;  APPLICANT:  Drmanac, Radoje T.
;  TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 784CIP2
;  CURRENT APPLICATION NUMBER: US/10/098,841
;  CURRENT FILING DATE:  2002-03-13
;  PRIOR APPLICATION NUMBER: 09/598,042
;  PRIOR FILING DATE: 2000-06-20
;  PRIOR APPLICATION NUMBER: 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  NUMBER OF SEQ ID NOS: 331
;  SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 83
;   LENGTH: 3262
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (121)..(2334)
US-10-098-841-83

Alignment Scores:
Pred. No.:                2.05e-141      Length:          3262
Score:                    1183.50        Matches:         230
Percent Similarity:       55.34%         Conservative:    55
Best Local Similarity:    44.66%         Mismatches:      65
Query Match:              61.23%         Indels:          165
DB:                       13             Gaps:            3

US-10-070-464-7 (1-360) x US-10-098-841-83 (1-3262)

Qy          2 GluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArgTyr 21
              :::|||   :::|||||||||||||||||||||||||:::||||||||||||||||||:::
Db        499 GATGACCCCAAGTCTGCGGGTGTGGCCACCTTCGTCATACAGGAAGAGTTCGACCGCTTC 558

Qy         22 SerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly---LysIleLeu 40
              :::||||||||||||||||||   |||           |||   |||   |||   |||
Db        559 ACTGGGTACTGGTGGTGCCCCACAGCCTCCTGGGAAGGTTCAGAGGGCCTCAAGACGCTG 618

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||   ||||||||||||||||||:::|||||||||   ||||||
Db        619 CGAATCCTGTATGAGGAAGTCGATGAGTCCGAGGTGGAGGTCATTCACGTCCCCTCTCCT 678

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
                ||||||   |||:::   ||||||:::|||||||||:::|||||||:::   ||||||
Db        679 GCGCTAGAAGAAAGGAAGACGGACTCGTATCGGTACCCCAGGACAGGCAGCAAGAATCCC 738

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              |||:::      |||:::::::|||          |||:::::::|||:::|||:::
Db        739 AAGATTGCCTTGAAACTGGCTGAGTTCCAGACTGACAGCCAGGGCAAGATCGTCTCGACC 798

Qy         98 ------------------------------------------------------------ 98

Db        799 CAGGAGAAGGAGCTGGTGCAGCCCTTCAGCTCGCTGTTCCCGAAGGTGGAGTACATCGCC 858

Qy         98 ------------------------------------------------------------ 98

Db        859 AGGGCCGGGTGGACCCGGGATGGCAAATACGCCTGGGCCATGTTCCTGGACCGGCCCCAG 918

Qy         98 ------------------------------------------------------------ 98

Db        919 CAGTGGCTCCAGCTCGTCCTCCTCCCCCCGGCCCTGTTCATCCCGAGCACAGAGAATGAG 978

Qy         98 ------------------------------------------------------------ 98

Db        979 GAGCAGCGGCTAGCCTCTGCCAGAGCTGTCCCCAGGAATGTCCAGCCGTATGTGGTGTAC 1038

Qy         98 ------------------------------------------------------------ 98

Db       1039 GAGGAGGTCACCAACGTCTGGATCAATGTTCATGACATCTTCTATCCCTTCCCCCAATCA 1098

Qy         98 ------------------------------------------------------------ 98

Db       1099 GAGGGAGAGGACGAGCTCTGCTTTCTCCGCGCCAATGAATGCAAGACCGGCTTCTGCCAT 1158

Qy         98 ------------------------------------------------------------ 98

Db       1159 TTGTACAAAGTCACCGCCGTTTTAAAATCCCAGGGCTACGATTGGAGTGAGCCCTTCAGC 1218

Qy         99 ----------------------------------------ValAspGluValArgArgLeu 105
                                                      |||:::|||      :::|||
Db       1219 CCCGGGGAAGGTGAGCAGAGCCTGACGAATGCTATCTGGGTCAATGAGGAGACCAAGCTG 1278

Qy        106 ValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyr 125
              |||||||||:::|||||||||||:::|||||||||||||||||||||||||||||||||||
Db       1279 GTGTACTTCCAGGGCACCAAGGACACGCCGCTGGAGCACCACCTCTACGTGGTCAGCTAT 1338

Qy        126 ValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIle 145
                       ||||||:::   |||||||||      |||:::||||||||||||   :::
Db       1339 GAGGCGGCCGGCGAGATCGTACGCCTCACCACGCCCGGCTTCTCCCATAGCTGCTCCATG 1398

Qy        146 SerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysVal 165
              ||||||:::   |||   |||:::|||   ||||||:::          |||   ||||||
Db       1399 AGCCAGAACTTCGACATGTTCGTCAGCCACTACAGCAGCGTGAGCACGCCGCCCTGCGTG 1458

Qy        166 SerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrp 185
                 :::||||||||||||   |||:::|||||||||      |||           ||||||
Db       1459 CACGTCTACAAGCTGAGCGGCCCCGACGACGACCCCCTGCACAAGCAGCCCCGCTTCTGG 1518

Qy        186 AlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSer 205
              |||:::::::::::::::|||                              :::||||||
Db       1519 GCTAGCATGATGGAGGCAGCC--------------------------AAGATCTTCCAT 1551

Qy        206 PheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGln 225
              |||   :::   :::          ||||||||||||:::|||||||||||   ||||||
Db       1552 TTCCACACGCGCTCGGATGTGCGGCTCTACGGCATGATCTACAAGCCCCACGCCTTGCAG 1611

Qy        226 ProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuVal 245
              ||||||||||||:::|||||||||||||:::|||||||||||||||||||||||||||||
Db       1612 CCAGGGAAGAAGCACCCCACCGTCCTCTTTGTATATGGAGGCCCCCAGGTGCAGCTGGTG 1671

Qy        246 AsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyr 265
              ||||||   ||||||||||:::||||||   ||||||||||||||||||||||||||||||
Db       1672 AATAACTCCTTCAAAGGCATCAAGTACTTGCGGCTCAACACACTGGCCTCCCTGGGCTAC 1731

Qy        266 ValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAla 285
              ||||||||||||||   ||||||||||||   |||||||||:::||||||||||||||
Db       1732 GCCGTGGTTGTGATTGACGGCAGGGGCTCCTGTCAGCGAGGGCTTCGGTTCGAAGGGGCC 1791

Qy        286 PheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeu 305
              |||   :::||||||||||:::||||||:::||||||||||||||||||||||||:::::
Db       1792 CTGAAAAACCAAATGGGCCAGGTGGAGATCGAGGACCAGGTGGAGGGCCTGCAGTTCGTG 1851

Qy        306 AlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGly 325
              |||   :::|||   ||||||||||||   ||||||   ||||||||||||||||||||
Db       1852 GCCGAGAAGTATGGCTTCATCGACCTGAGCCGAGTTGCCATCCATGGCTGGTCCTACGGG 1911

Qy        326 GlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAla 345
              |||:::||||||||||||   |||:::   :::      :::|||:::||||||||||||
Db       1912 GGCTTCCTCTCGCTCATGGGGCTAATCCACAAGCCCCAGGTGTTCAAGGTGGCCATCGCG 1971

Qy        346 GlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              |||||||||||||:::|||:::   ||||||||||||||||||
Db       1972 GGTGCCCCGGTCACCGTCTGGATGGCCTACGACACAGGGTACACT 2016

Search completed: October 16, 2003, 09:10:05
Job time : 309.152 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:24:12 ; Search time 2162.86 Seconds
(without alignments)
4045.398 Million cell updates/sec

Title: US-10-070-464-7
Perfect score: 1933
Sequence: 1 EEDARSAGVATFVLQEEFDR.....RVAIAGAPVTLWIFYDTGYT 360

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10070464/runat_15102003_113553_24829/app_query.fasta_1.2652
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10070464_@CGN_1_1_8056_@runat_15102003_113553_24829 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1756.5	90.9	3143	11	AK016546	AK016546 Mus muscu
2	1669	86.3	5517	11	AK029788	AK029788 Mus muscu
3	1415	73.2	1042	12	BM557438	BM557438 AGENCOURT
4	1325.5	68.6	910	13	BQ675260	BQ675260 AGENCOURT
5	1297	67.1	957	13	BQ675006	BQ675006 AGENCOURT
6	1237.5	64.0	1041	12	BI084090	BI084090 602869453
7	1232	63.7	746	14	BY751026	BY751026 BY751026
8	1205	62.3	968	13	BQ671635	BQ671635 AGENCOURT
9	1204.5	62.3	3327	11	AK050021	AK050021 Mus muscu
10	1204.5	62.3	3376	11	AK050023	AK050023 Mus muscu
11	1204.5	62.3	3457	11	AK078301	AK078301 Mus muscu
12	1122	58.0	1082	13	BU239476	BU239476 603322338
13	1117	57.8	674	10	BE888665	BE888665 601513061
14	1108.5	57.3	1090	12	BM910838	BM910838 AGENCOURT
15	1099	56.9	639	10	BG165205	BG165205 602344074
16	1089	56.3	753	9	AL040398	AL040398 DKFZp434A
17	1057	54.7	658	10	BE388695	BE388695 601286345
18	991	51.3	831	12	BG974324	BG974324 602844075
19	976.5	50.5	853	10	BF694553	BF694553 602081767
20	970.5	50.2	898	13	BU908783	BU908783 AGENCOURT
21	951.5	49.2	603	13	BU483718	BU483718 603470021
22	949	49.1	789	10	BG709118	BG709118 602675382
23	937	48.5	982	13	BQ678015	BQ678015 AGENCOURT
24	922	47.7	897	13	BU191638	BU191638 AGENCOURT
25	919	47.5	695	10	BG715866	BG715866 602675282
26	916.5	47.4	853	12	BI223892	BI223892 602941035
27	905	46.8	820	13	BU522948	BU522948 AGENCOURT
28	902.5	46.7	845	13	BU216915	BU216915 603109373
29	887.5	45.9	896	12	BG921463	BG921463 602824365
30	868	44.9	788	10	BG163397	BG163397 602338360
31	866	44.8	871	13	BU196183	BU196183 AGENCOURT
32	863	44.6	821	12	BI690774	BI690774 603314392
33	833	43.1	931	13	BU235022	BU235022 603409569
34	826	42.7	479	12	BI135177	BI135177 UI-M-BH3-
35	788.5	40.8	1021	12	BG974587	BG974587 602844820
c 36	787	40.7	687	14	CB851806	CB851806 UI-CF-FN0
37	765	39.6	713	14	CA321029	CA321029 UI-M-FW0-
38	762	39.4	432	13	BY520490	BY520490 BY520490
39	762	39.4	957	13	BU234147	BU234147 603412033
40	756	39.1	741	14	CB595102	CB595102 AGENCOURT
41	753	39.0	551	9	AL643656	AL643656 AL643656
42	746	38.6	735	13	BU447277	BU447277 603766685
43	738	38.2	428	13	BY548432	BY548432 BY548432
44	736	38.1	1023	13	BQ681703	BQ681703 AGENCOURT
45	735	38.0	1080	10	BF203783	BF203783 601868764

ALIGNMENTS

RESULT 1
AK016546
LOCUS AK016546 3143 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4932434F09 product:DIPEPTIDYL PEPTIDASE 8 homolog [Homo sapiens], full insert sequence.
ACCESSION AK016546
VERSION AK016546.1 GI:12855334
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 3143)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised longer than 7 kb was selected before cloning. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.
FEATURES Location/Qualifiers
source 1. .3143
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:4932434F09"
/db_xref="MGI:1907554"
/db_xref="taxon:10090"
/clone="4932434F09"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
CDS 327. .3005
/note="unnamed protein product; DIPEPTIDYL PEPTIDASE 8 homolog [Homo sapiens] (SPTR|Q9HBM5, evidence: FASTY, 95.7%ID, 100%length, match=2649)
putative"
/codon_start=1
/protein_id="BAB30295.2"
/db_xref="GI:26382128"
/db_xref="MGI:1921638"
/translation="MKIPSGRCNMAAAMETEQLGVEIFETAECEEGNGESQDRPKLEP
FYVERYSWSQLKKLLADTRKYHGYMMAKAPHDFMFVKRTDPDRPHSDRVYYLAMSGEN
RENTLFYSEIPKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGI
AAYDYHPGSGTFLFQAGSGIYHIKDGGPHGFTQQPLRPNLVETSCPNIRMDPKLCPAD
PDWIAFIHSNDIWISNLVTREERRITYVHNELANMEEDPRSAGVATFVLQEEFDRYSG
YWWCPQAERTPSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPK
VTFKMSEIVVDAAGGIIDVIDKELVQPPEILFEGVEYIARAGWTPEGKHAWSILLDRS
QTHLQIVLISPELFIPVEDDAMDRQRLIESVPPSVTPLIIYEETTDIWINIHDIFHVF
PQTHEDEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEITI
TSGEWEVLGRHGSNIWVDEARKLVYFEGTKDSPLEHHLYVTSYANPGEVVRLTDRGYS
HSCCLSRHCDFFISKYSNQKNPHCVSLYKLSSPEDDPVHKTKEFWATILDSAGPLPDY
TPPEIFSPESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFR
LNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASQYDFIDL
DRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQ
GYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQ
ERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI"
BASE COUNT 894 a 688 c 773 g 788 t
ORIGIN

Alignment Scores:

Pred. No.:	1.23e-190	Length:	3143
Score:	1756.50	Matches:	343
Percent Similarity:	65.67%	Conservative:	7
Best Local Similarity:	64.35%	Mismatches:	10
Query Match:	90.87%	Indels:	173
DB:	11	Gaps:	1

US-10-070-464-7 (1-360) x AK016546 (1-3143)

Qy 1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20

```
Db       1086 GAAGAGGATCCCAGATCAGCTGGTGTTGCCACCTTTGTTCTTCAAGAAGAATTTGACAGA 1145

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40

Db       1146 TACTCTGGCTACTGGTGGTGTCCCCAAGCGGAAAGAACTCCTAGTGGTGGTAAAATTCTT 1205

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60

Db       1206 AGAATTCTCTATGAAGAAAATGATGAATCTGAGGTGGAGATTATTCATGTTACGTCCCCC 1265

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80

Db       1266 ATGTTGGAAACAAGGAGGGCAGATTCCTTTCGTTATCCCAAAACAGGCACAGCAAACCCA 1325

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98

Db       1326 AAGGTCACTTTCAAGATGTCGGAGATTGTTGTTGATGCTGCAGGAGGGATTATAGATGTC 1385

Qy         98 ------------------------------------------------------------ 98

Db       1386 ATAGATAAAGAACTGGTTCAACCTTTCGAGATTCTGTTTGAGGGAGTTGAATATATTGCC 1445

Qy         98 ------------------------------------------------------------ 98

Db       1446 AGAGCCGGATGGACTCCAGAGGGAAAACATGCCTGGTCCATCCTACTAGACCGTTCCCAG 1505

Qy         98 ------------------------------------------------------------ 98

Db       1506 ACTCACCTGCAGATAGTTCTGATCTCCCCTGAGTTATTCATCCCAGTAGAAGATGATGCC 1565

Qy         98 ------------------------------------------------------------ 98

Db       1566 ATGGACAGACAGAGACTTATAGAGTCAGTTCCTGACTCTGTGACACCACTGATCATCTAT 1625

Qy         98 ------------------------------------------------------------ 98

Db       1626 GAAGAAACAACAGACATCTGGATAAATATCCACGATATTTTTCATGTTTTTCCTCAAACT 1685

Qy         98 ------------------------------------------------------------ 98

Db       1686 CATGAAGATGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTTCGTCATCTG 1745

Qy         98 ------------------------------------------------------------ 98

Db       1746 TATAAAATCACATCCATTTTAAAGGAGAGCAAATATAAACGGTCCAGTGGTGGACTACCT 1805

Qy         98 ------------------------------------------------------------ 98

Db       1806 GCCCCAAGTGATTTCAAGTGTCCTATCAAAGAAGAAATAACAATTACCAGTGGTGAATGG 1865

Qy         99 ---------------------------------ValAspGluValArgArgLeuValTyr 107

Db       1866 GAAGTACTTGGCCGGCATGGATCTAATATCTGGGTTGATGAAGCCAGAAAGCTGGTGTAC 1925

Qy        108 PheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsn 127

Db       1926 TTTGAAGGCACCAAAGACTCTCCTTTGGAGCATCACCTGTACGTGACCAGTTATGCAAAC 1985

Qy        128 ProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGln 147

Db       1986 CCTGGAGAAGTGGTGAGGCTGACTGACCGTGGCTACTCACACTCCTGCTGCCTCAGCCGG 2045

Qy        148 HisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeu 167

Db       2046 CATTGTGACTTCTTCATAAGTAAGTACAGCAACCAGAAGAATCCACACTGTGTGTCCCTC 2105

Qy        168 TyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThr 187

Db       2106 TACAAACTCTCAAGTCCTGAGGATGACCCAGTTCATAAAACAAAGGAATTTTGGGCCACC 2165

Qy        188 IleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGlu 207

Db       2166 ATTTTGGATTCAGCAGGTCCTCTTCCTGACTACACCCCTCCAGAAATTTTTTCTTTTGAA 2225

Qy        208 SerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGly 227

Db       2226 AGTACTACTGGATTTACACTGTATGGAATGTTGTATAAGCCTCATGACCTACAACCTGGA 2285

Qy        228 LysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsn 247

Db       2286 AAGAAATACCCCACTGTGTTATTCATATATGGTGGTCCCCAGGTGCAGCTGGTGAACAAT 2345

Qy        248 ArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValVal 267

Db       2346 CGGTTTAAAGGAGTCAAGTATTTCCGCCTGAACACCCTGGCCTCCCTGGGTTATGTGGTT 2405

Qy        268 ValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLys 287

Db       2406 GTGGTGATAGACAACAGGGGATCCTGTCACCGAGGACTTAAATTTGAAGGCGCCTTTAAA 2465

Qy        288 TyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSer 307

Db       2466 TATAAAATGGGTCAAATAGAAATCGATGATCAAGTGGAAGGACTCCAGTACCTAGCATCT 2525

Qy        308 ArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyr 327

Db       2526 CAGTATGACTTCATTGACTTGGATCGAGTGGGCATCCACGGCTGGTCCTATGGTGGCTAC 2585

Qy        328 LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla 347

Db       2586 CTCTGCCTGATGGCATTAATGCAGAGGTCGGATATCTTCCGGGTGGCTATTGCTGGGGCC 2645

Qy        348 ProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360

Db       2646 CCAGTCACCCTGTGGATCTTCTATGATACAGGATACACG 2684
```

RESULT 2
AK029788

```
LOCUS       AK029788               5517 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus adult male testis cDNA, RIKEN full-length enriched
            library, clone:4930560C15 product:DIPEPTIDYL PEPTIDASE 8 homolog
            [Homo sapiens], full insert sequence.
ACCESSION   AK029788
VERSION     AK029788.1  GI:26081520
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
```

PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 5517)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.
FEATURES Location/Qualifiers
source 1. .5517
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:4930560C15"
/db_xref="taxon:10090"
/clone="4930560C15"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
misc_feature 2221. .4837
/note="DIPEPTIDYL PEPTIDASE 8 homolog [Homo sapiens] (SPTR|Q9HBM5, evidence: FASTY, 95.7%ID, 100%length, match=2649)
putative"
BASE COUNT 1514 a 1224 c 1230 g 1549 t
ORIGIN

Alignment Scores:

Pred. No.:	3.55e-180	Length:	5517
Score:	1669.00	Matches:	332
Percent Similarity:	66.67%	Conservative:	10
Best Local Similarity:	64.72%	Mismatches:	18
Query Match:	86.34%	Indels:	153
DB:	11	Gaps:	4

US-10-070-464-7 (1-360) x AK029788 (1-5517)

```
Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2980 GAAGAGGATCCCAGATCAGCTGGTGTTGCCACCTTTGTTCTTCAAGAAGAATTTGACAGA 3039

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||:::||||||   ||||||||||||||||||||||||||
Db       3040 TACTCTGGCTACTGGTGGTGTCCCCAAGCGGAAAGAACTCCTAGTGGTGGTAAAATTCTT 3099

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3100 AGAATTCTCTATGAAGAAAATGATGAATCTGAGGTGGAGATTATTCATGTTACGTCCCCC 3159

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLys---------------------- 74
              ||||||||||||||||||||||||||||||||||||||||||
Db       3160 ATGTTGGAAACAAGGAGGGCAGATTCCTTTCGTTATCCCAAAACAGATTATAGATGTCAT 3219

Qy         75 ---------ThrGlyThrAla---------------------------------------- 78
                      |||||:::
Db       3220 AGATAAAGAACTGGTTCAACCTTTCGAGATTCTGTTTGAGGGAGTTGAATATATTGCCAG 3279

Qy         78 ------------------------------------------------------------ 78

Db       3280 AGCCGGATGGACTCCAGAGGGAAAACATGCCTGGTCCATCCTACTAGACCGTTCCCAGAC 3339

Qy         79 AsnPro-LysValThrPheLysMetSerGluIleMetIleAspAlaGlu----------- 94
              :::||| ::::::          ||:::   |||     |||
Db       3340 TCACCTGCAGATAGTTCTGATCTCCCCTGAGTTATTCATCCCAGTAGAAGATGATGCCAT 3399

Qy         94 ------------------------------------------------------------ 94

Db       3400 GGACAGACAGAGACTTATAGAGTCAGTTCCTGACTCTGTGACACCACTGATCATCTATGA 3459

Qy         94 ------------------------------------------------------------ 94

Db       3460 AGAAACAACAGACATCTGGATAAATATCCACGATATTTTTCATGTTTTTCCTCAAACTCA 3519

Qy         94 ------------------------------------------------------------ 94

Db       3520 TGAAGATGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTTCGTCATCTGTA 3579

Qy         94 ------------------------------------------------------------ 94

Db       3580 TAAAATCACATCCATTTTAAAGGAGAGCAAATATAAACGGTCCAGTGGTGGACTACCTGC 3639

Qy         94 ------------------------------------------------------------ 94

Db       3640 CCCAAGTGATTTCAAGTGTCCTATCAAAGAAGAAATAACAATTACCAGTGGTGAATGGGA 3699

Qy         95 -------GlyArg------------IleIleValAspGluValArgArgLeuValTyrPh 108
                     ||||||            |||   |||||||||   ||:::||||||||||
Db       3700 AGTACTTGGCCGGCATGGATCTAATATCTGGGTTGATGAAGCCAGAAAGCTGGTGTACTT 3759

Qy        108 eGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrValAsnPr 128
              ||||||||||||||||||||||||||||||||||||||||||||||||   ||||||   |||||
Db       3760 TGAAGGCACCAAAGACTCTCCTTTGGAGCATCACCTGTACGTGACCAGTTATGCAAACCC 3819
```

```
Qy        128 oGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSerGlnHi 148
              ||||||||||  |||||||||||||||||||||||||||||||:::|||:::||
Db       3820 TGGAGAAGTGGTGAGGCTGACTGACCGTGGCTACTCACACTCCTGCTGCCTCAGCCGGCA 3879

Qy        148 sCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSerLeuTy 168
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3880 TTGTGACTTCTTCATAAGTAAGTACAGCAACCAGAAGAATCCACACTGTGTGTCCCTCTA 3939

Qy        168 rLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAlaThrIl 188
              |||||||||||||||||||||||||||      |||||||||||||||||||||||||
Db       3940 CAAACTCTCAAGTCCTGAGGATGACCCAGTTCATAAAACAAAGGAATTTTGGGCCACCAT 3999

Qy        188 eLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPheGluSe 208
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4000 TTTGGATTCAGCAGGTCCTCTTCCTGACTACACCCCTCCAGAAATTTTTTCTTTTGAAAG 4059

Qy        208 rThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnProGlyLy 228
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4060 TACTACTGGATTTACACTGTATGGAATGTTGTATAAGCCTCATGACCTACAACCTGGAAA 4119

Qy        228 sLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsnAsnAr 248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4120 GAAATACCCCACTGTGTTATTCATATATGGTGGTCCCCAGGTGCAGCTGGTGAACAATCG 4179

Qy        248 gPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrValValVa 268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4180 GTTTAAAGGAGTCAAGTATTTCCGCCTGAACACCCTGGCCTCCCTGGGTTATGTGGTTGT 4239

Qy        268 lValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPheLysTy 288
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4240 GGTGATAGACAACAGGGGATCCTGTCACCGAGGACTTAAATTTGAAGGCGCCTTTAAATA 4299

Qy        288 rLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAlaSerAr 308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||::
Db       4300 TAAAATGGGTCAAATAGAAATCGATGATCAAGTGGAAGGACTCCAGTACCTAGCATCTCA 4359

Qy        308 gTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyrLe 328
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4360 GTATGACTTCATTGACTTGGATCGAGTGGGCATCCACGGCTGGTCCTATGGTGGCTACCT 4419

Qy        328 uSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAlaPr 348
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4420 CTCCCTGATGGCATTAATGCAGAGGTCGGATATCTTCCGGGTGGCTATTGCTGGGGCCCC 4479

Qy        348 oValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||||||||||||||||||||||||||
Db       4480 AGTCACCCTGTGGATCTTCTATGATACAGGATACACG 4516
```

RESULT 3
BM557438
LOCUS BM557438 1042 bp mRNA linear EST 20-FEB-2002
DEFINITION AGENCOURT_6578992 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5466759 5', mRNA sequence.
ACCESSION BM557438
VERSION BM557438.1 GI:18799430
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1042)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLCM1967 row: k column: 16
High quality sequence stop: 697.
FEATURES Location/Qualifiers
source 1. .1042
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5466759"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_41"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 279 a 223 c 234 g 303 t 3 others
ORIGIN

Alignment Scores:

Pred. No.:	4.97e-152	Length:	1042
Score:	1415.00	Matches:	262
Percent Similarity:	99.24%	Conservative:	0
Best Local Similarity:	99.24%	Mismatches:	2
Query Match:	73.20%	Indels:	0
DB:	12	Gaps:	0

US-10-070-464-7 (1-360) x BM557438 (1-1042)

```
Qy         97 IleIleValAspGluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeu 116
              |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         28 ATCCAAGTTGATGAAGTCAGAAGGCTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTA 87

Qy        117 GluHisHisLeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAsp 136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         88 GAGCATCACCTGTACGTAGTCAGTTACGTAAATCCTGGAGAGGTGACAAGGCTGACTGAC 147

Qy        137 ArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyr 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        148 CGTGGCTACTCACATTCTTGCTGCATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTAT 207

Qy        157 SerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAsp 176
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 AGTAACCAGAAGAATCCACACTGTGTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGAC 267

Qy        177 ProThrCysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuPro 196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        268 CCAACTTGCAAAACAAAGGAATTTTGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCT 327

Qy        197 AspTyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGly 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        328 GACTATACTCCTCCAGAAATTTTCTCTTTTGAAAGTACTACTGGATTTACATTGTATGGG 387

Qy        217 MetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 236
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        388 ATGCTCTACAAGCCTCATGATCTACAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATA 447

Qy        237 TyrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArg 256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        448 TATGGTGGTCCTCAGGTGCAGTTGGTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGC 507

Qy        257 LeuAsnThrLeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCys 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        508 TTGAATACCCTAGCCTCTCTAGGTTATGTGGTTGTAGTGATAGACAACAGGGGATCCTGT 567

Qy        277 HisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAsp 296
              |||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db        568 CACCGAGGGCTTAAATTTGAAGGCGCCTTTANATATAAAATGGGTCAAATAGAAATTGAC 627

Qy        297 AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 316
```

```
Db        628 GATCAGGTGGAAGGACTCCAATATCTAGCTTCTCGATATGATTTCATTGACTTAGATCGT 687

Qy        317 ValGlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 336

Db        688 GTGGGCATCCACGGCTGGTCCTATGGAGGATACCTCTCCCTGATGGCATTAATGCAGAGG 747

Qy        337 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuTrpIlePheTyrAsp 356

Db        748 TCAGATATCTTCAGGGTTGCTATTGCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGAT 807

Qy        357 ThrGlyTyrThr 360

Db        808 ACAGGATACACG 819

RESULT 4
BQ675260
LOCUS       BQ675260                910 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8354972 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6275478
            5', mRNA sequence.
ACCESSION   BQ675260
VERSION     BQ675260.1  GI:21786094
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 910)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2458  row: 1  column: 07
            High quality sequence stop: 618.
FEATURES             Location/Qualifiers
     source          1. .910
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6275478"
                     /tissue_type="epidermoid carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_102"
                     /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of California, Berkeley) using ZAP-cDNA
                     synthesis kit (Stratagene) and Superscript II RT (Life
                     Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      248 a    186 c    212 g    262 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:              8.28e-142      Length:          910
Score:                  1325.50        Matches:         254
Percent Similarity:     96.24%         Conservative:    2
Best Local Similarity:  95.49%         Mismatches:      6
Query Match:            68.57%         Indels:          4
DB:                     13             Gaps:            1

US-10-070-464-7 (1-360) x BQ675260 (1-910)

Qy         97 IleIleValAspGluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeu 116

Db        116 ATCCAAGTTGATGAAGTCAGAAGGCTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTA 175

Qy        117 GluHisHisLeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAsp 136

Db        176 GAGCATCACCTGTACGTAGTCAGTTACGTAAATCCTGGAGAGGTGACAAGGCTGACTGAC 235

Qy        137 ArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyr 156

Db        236 CGTGGCTACTCACATTCTTGCTGCATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTAT 295

Qy        157 SerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAsp 176

Db        296 AGTAACCAGAAGAATCCACACTGTGTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGAC 355

Qy        177 ProThrCysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuPro 196

Db        356 CCAACTTGCAAAACAAAGGAATTTTGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCT 415

Qy        197 AspTyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGly 216

Db        416 GACTATACTCCTCCAGAAATTTTCTCTTTTGAAAGTACTACTGGATTTACATTGTATGGG 475

Qy        217 MetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 236

Db        476 ATGCTCTACAAGCCTCATGATCTACAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATA 535

Qy        237 TyrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArg 256

Db        536 TATGGTGGTCCTCAGGTGCAGTTGGTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGC 595

Qy        257 LeuAsnThrLeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCys 276

Db        596 TTGAATACCCTAGCCTCTCTAGGTTATGTGGTTGTAGTGATAGACAACAGGGGATCCTGT 655

Qy        277 HisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAsp 296

Db        656 CACCGAGGGCTTAAATTTGAAGGCGCCTTTAAATATAAAATGGGTCNAATAGAAAATGAC 715

Qy        297 AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 316

Db        716 GATCAGGTGGAAGGACTCCAATATCTAGCTTCTCGATATGATTTCATTGACTTAGATCGT 775

Qy        317 ValGlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAla-LeuMetGlnAr 336

Db        776 GTGGGCATCCACGGCTGGTCCTATGGAGGATACCTCTCCCTGATGGCATTTAATGCAGAA 835

Qy        336 gSerAsp-IlePheArgValAlaIleAlaGlyAlaProValThrLeuTrp---Ile-Phe 354

Db        836 GTCAGATTATCTTCAAGGTTGCTATTGCTGGGGCCCCAGNTCAATTCTGGGGGATCTTTC 895

Qy        355 TyrAspThrGly 358

Db        896 TATGATACAGGG 907

RESULT 5
BQ675006
LOCUS       BQ675006                957 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8303592 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6274622
            5', mRNA sequence.
ACCESSION   BQ675006
VERSION     BQ675006.1  GI:21785840
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 957)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2456  row: h  column: 15
            High quality sequence start: 4
            High quality sequence stop: 624.
FEATURES             Location/Qualifiers
     source          1. .957
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6274622"
                     /tissue_type="epidermoid carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_102"
                     /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of  California, Berkeley) using ZAP-cDNA
                     synthesis kit (Stratagene) and Superscript II RT (Life
                     Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      257 a    206 c    228 g    266 t
ORIGIN

Alignment Scores:
Pred. No.:              1.72e-138       Length:         957
Score:                  1297.00         Matches:        253
Percent Similarity:     95.15%          Conservative:   2
Best Local Similarity:  94.40%          Mismatches:     8
Query Match:            67.10%          Indels:         5
DB:                     13              Gaps:           0

US-10-070-464-7 (1-360) x BQ675006 (1-957)

Qy          97 IleIleValAspGluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeu 116
               |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         129 ATCCAAGTTGATGAAGTCAGAAGGCTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTA 188

Qy         117 GluHisHisLeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAsp 136
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         189 GAGCATCACCTGTACGTAGTCAGTTACGTAAATCCTGGAGAGGTGACAAGGCTGACTGAC 248

Qy         137 ArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyr 156
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         249 CGTGGCTACTCACATTCTTGCTGCATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTAT 308

Qy         157 SerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAsp 176
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         309 AGTAACCAGAAGAATCCACACTGTGTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGAC 368

Qy         177 ProThrCysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuPro 196
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         369 CCAACTTGCAAAACAAAGGAATTTTGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCT 428

Qy         197 AspTyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGly 216
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         429 GACTATACTCCTCCAGAAATTTTCTCTTTTGAAAGTACTACTGGATTTACATTGTATGGG 488

Qy         217 MetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 236
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         489 ATGCTCTACAAGCCTCATGATCTACAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATA 548

Qy         237 TyrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArg 256
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         549 TATGGTGGTCCTCAGGTGCAGTTGGTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGC 608

Qy         257 LeuAsnThr-LeuAlaSerLeuGlyTyrValValValValIleAsp-AsnArgGlySerC 276
               |||||||| |||||||||||||||||||||||||||||||||||| |||||||||||||
Db         609 TTGAATACCCCTAGCCTCTCTAGGTTATGTGGTTGTAGTGATAGACAAACAGGGGATCCT 668

Qy         276 ysHisArgGlyLeuLysPheGluGlyAla-PheLysTyrLysMet-GlyGlnIleGluIl 295
               |||||||||||||||||||||||||||| |||  |||||||| ||||||||||||||||
Db         669 GTCACCGAGGGCTTAAATTTGAAGGCGCCCTTTTAATATAAAATGGGGTCAAATAGAAAT 728

Qy         295 eAspAspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAs 315
               ||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db         729 TGACGATCAGGGGGAAGGACTCCAATATCTAGCTTCTCGATATGATTTCATTGACTTAGA 788

Qy         315 pArgValGlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGl 335
               |||||||||||||||||||||||||||||||  |||||||||  |||  |||||||||||
Db         789 TCGTGTGGGCATCCACGGCTGGTCCTATGGAGGATCCCTCTCCCTGAAGGCATTTATGCA 848

Qy         335 nArgSerAspIlePheArgValAlaIleAlaGlyAlaProValThr-LeuTrpIlePheT 355
               ||||||||||||||:::|||||||||||||||  |||||||||  |||||||||||||||:
Db         849 GAGGTCAGATATCTTCCAGGTTGCTATTGCTGGGGGCCCAGTCACTCCGTTGGATCTTCC 908

Qy         355 yrAspThrGlyTyr 359
               ::|||||||||||||
Db         909 ATGATACAGGATAC 922

RESULT 6
BI084090
LOCUS       BI084090                1041 bp    mRNA    linear   EST 20-JUN-2001
DEFINITION  602869453F1 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:5013996 5',
            mRNA sequence.
ACCESSION   BI084090
VERSION     BI084090.1  GI:14502420
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1820  row: j  column: 13
            High quality sequence stop: 847.
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5013996"
                     /tissue_type="epidermoid carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_102"
                     /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of  California, Berkeley) using ZAP-cDNA
                     synthesis kit (Stratagene) and Superscript II RT (Life
                     Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      280 a    227 c    242 g    292 t
ORIGIN

Alignment Scores:
Pred. No.:              1.41e-131       Length:         1041
Score:                  1237.50         Matches:        254
```

```
Percent Similarity:     93.77%          Conservative:  2
Best Local Similarity:  93.04%          Mismatches:    8
Query Match:            64.02%          Indels:        11
DB:                     12              Gaps:          0

US-10-070-464-7 (1-360) x BI084090 (1-1041)

Qy          97 IleIleValAspGluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeu 116
               |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         120 ATCCAAGTTGATGAAGTCAGAAGGCTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTA 179

Qy         117 GluHisHisLeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAsp 136
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         180 GAGCATCACCTGTACGTAGTCAGTTACGTAAATCCTGGAGAGGTGACAAGGCTGACTGAC 239

Qy         137 ArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyr 156
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         240 CGTGGCTACTCACATTCTTGCTGCATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTAT 299

Qy         157 SerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAsp 176
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         300 AGTAACCAGAAGAATCCACACTGTGTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGAC 359

Qy         177 ProThrCysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuPro 196
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         360 CCAACTTGCAAAACAAAGGAATTTTGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCT 419

Qy         197 AspTyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGly 216
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         420 GACTATACTCCTCCAGAAATTTTCTCTTTTGAAAGTACTACTGGATTTACATTGTATGGG 479

Qy         217 MetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 236
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         480 ATGCTCTACAAGCCTCATGATCTACAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATA 539

Qy         237 TyrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGly--ValLysTyrPheA 256
               ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db         540 TATGGTGGTCCTCAGGTGCAGTTGGTGAATAATCGGTTTAACAGCGAGTCAAGTATTTCC 599

Qy         256 rgLeuAsnThrLeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerC 276
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         600 GCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTTGTAGTGATAGACAACAGGGGATCCT 659

Qy         276 ysHisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleA 296
               |||||||||||||||||||||||||||||||  |||  |||||||||  |||||||||||
Db         660 GTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAA-TATACAATGGGTCAATA-GAAATTG 717

Qy         296 spAspGlnValGluGlyLeuGln-TyrLeuAlaSerArgTyrAspPheIle-AspLeuAs 315
               |||||   ||||||||||||||| ||||||||||||||||||||||||| ||||||||||
Db         718 ACGATCACGTGGAAGGACTCCAATTATCTAGCTTCTCGATATGATTTCATTTGACTTAGA 777

Qy         315 pArgValGlyIleHisGlyTrpSerTyrGly-GlyTyrLeuSerLeu-MetAlaLeuMet 334
               |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||||
Db         778 TCGTGTGGGCATCCACGGGTGGTCCTATGGAGGGATACCTCTCCCTGAATGGCATTAATG 837

Qy         335 GlnArgSerAsp-IlePheArgValAlaIleAla-GlyAlaProValThrLeuTrpIleP 354
               |||||||||||| :::|||:::   |||||||||| ||||||||||||||||||||||||
Db         838 CAGAGGTCAGATACTCTTCCAGGGTGCAATAGCTGGGGGCCCCAGTCACTCTGTGGATCT 897

Qy         354 he-TyrAspThrGlyTyrThr 360
               || ||||||||||||||||||
Db         898 TCTTATGATACAGGATACACG 918

RESULT 7
BY751026
LOCUS       BY751026                 746 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  BY751026 RIKEN full-length enriched, activated spleen Mus musculus
            cDNA clone F830018D07 5', mRNA sequence.
ACCESSION   BY751026
VERSION     BY751026.1  GI:27181193
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 746)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
            ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
            Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
            ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
            Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
            Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
            Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
            Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
            Trust/MRC building Addenbrookes Hospital Cambridge) whose
            assistance we gratefully acknowledge.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
```

```
FEATURES             Location/Qualifiers
     source          1. .746
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NOD"
                     /db_xref="taxon:10090"
                     /clone="F830018D07"
                     /tissue_type="activated spleen"
                     /clone_lib="RIKEN full-length enriched, activated spleen"
BASE COUNT      194 a    178 c    176 g    197 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              3.62e-131       Length:         746
Score:                  1232.00         Matches:        229
Percent Similarity:     97.48%          Conservative:   3
Best Local Similarity:  96.22%          Mismatches:     5
Query Match:            63.74%          Indels:         1
DB:                     14              Gaps:           0

US-10-070-464-7 (1-360) x BY751026 (1-746)

Qy        124 SerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCys 143
              ||||||   ||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db          1 AGTTATGCAAACCCTGGAGAAGTGGTGAGGCTGACTGACCGTGGCTACTCACACTCCTGC 60

Qy        144 CysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHis 163
              |||:::||:::||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGCCTCAGCCGGCATTGTGACTTCTTCATAAGTAAGTACAGCAACCAGAAGAATCCACAC 120

Qy        164 CysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGlu 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGTGTGTCCCTCTACAAACTCTCAAGTCCTGAGGATGACCCAGTTCATAAAACAAAGGAA 180

Qy        184 PheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIle 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTTGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTACACCCCTCCAGAAATT 240

Qy        204 PheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAsp 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTTTCTTTTGAAAGTACTACTGGATTTACACTGTATGGAATGTTGTATAAGCCTCATGAC 300

Qy        224 LeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTACAACCTGGAAAGAAATACCCCACTGTGTTATTCATATATGGTGGTCCCCAGGTGCAG 360

Qy        244 LeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeu 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGGTGAACAATCGGTTTAAAGGAGTCAAGTATTTCCGCCTGAACACCCTGGCCTCCCTG 420

Qy        264 GlyTyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGlu 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGTTATGTGGTTGTGGTGATAGACAACAGGGGATCCTGTCACCGAGGACTTAAATTTGAA 480

Qy        284 GlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGln 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGCGCCTTTAAATATAAAATGGGTCAAATAGAAATCGATGATCAAGTGGAAGGACTCCAG 540

Qy        304 TyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSer 323
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        541 TACCTAGCATCTCAGTATGACTTCATTGACTTGGATCGAGTGGGCATCCACGGCTGGTCC 600

Qy        324 TyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIle-PheArgValAl 343
              ||||||||||||||||||||||||||||||||||||||||||||||    |||||||||
Db        601 TATGGTGGCTACCTCTCCCTGATGGCATTAATGCAGAGGTCGGATATCTNTCCGGGTGGC 660

Qy        343 aIleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TATTGCTGGGGCCCCAGTCACCCTGTGGATCTTCTATGATACAGGATACACG 712

RESULT 8
BQ671635
LOCUS       BQ671635                968 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8203887 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6255645
            5', mRNA sequence.
ACCESSION   BQ671635
VERSION     BQ671635.1  GI:21782469
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 968)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2407  row: a  column: 22
            High quality sequence stop: 630.
FEATURES             Location/Qualifiers
     source          1. .968
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6255645"
                     /tissue_type="epidermoid carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_102"
                     /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of  California, Berkeley) using ZAP-cDNA
                     synthesis kit (Stratagene) and Superscript II RT (Life
                     Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      245 a    228 c    235 g    259 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              7e-128          Length:         968
Score:                  1205.00         Matches:        229
Percent Similarity:     97.46%          Conservative:   1
Best Local Similarity:  97.03%          Mismatches:     3
Query Match:            62.34%          Indels:         3
DB:                     13              Gaps:           0

US-10-070-464-7 (1-360) x BQ671635 (1-968)

Qy         97 IleIleValAspGluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeu 116
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        117 ATCCAAGTTGATGAAGTCAGAAGGCTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTA 176

Qy        117 GluHisHisLeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAsp 136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        177 GAGCATCACCTGTACGTAGTCAGTTACGTAAATCCTGGAGAGGTGACAAGGCTGACTGAC 236

Qy        137 ArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyr 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 CGTGGCTACTCACATTCTTGCTGCATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTAT 296

Qy        157 SerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAsp 176
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        297 AGTAACCAGAAGAATCCACACTGTGTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGAC 356

Qy        177 ProThrCysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuPro 196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        357 CCAACTTGCAAAACAAAGGAATTTTGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCT 416
```

```
Qy        197 AspTyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGly 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        417 GACTATACTCCTCCAGAAATTTTCTCTTTTGAAAGTACTACTGGATTTACATTGTATGGG 476

Qy        217 MetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 236
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        477 ATGCTCTACAAGCCTCATGATCTACAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATA 536

Qy        237 TyrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArg 256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        537 TATGGTGGTCCTCAGGTGCAGTTGGTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGC 596

Qy        257 LeuAsnThrLeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCys 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        597 TTGAATACCCTAGCCTCTCTAGGTTATGTGGTTGTAGTGATAGACAACAGGGGATCCTGT 656

Qy        277 HisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAsp 296
              |||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db        657 CACCGAGGGCTTAAATTTGAAGGCGCCTTTANATATAAAATGGGTCAAATAGAAATTGAC 716

Qy        297 AspGlnValGluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 316
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 GATCAGGTGGAAGGACTCCAATATCTAGCTTCTCGATATGATTTCATTGACTTAGATCGG 776

Qy        317 -ValGlyIleHisGlyTrpSer-Tyr-GlyGlyTyrLeuSer 329
               ||||||||||||   |||::: ||| ||||||||||||||
Db        777 TGTGGGCATCCACCGCTGGGCCCTATGGGAGGATACCTCTCC 818
```

RESULT 9
AK050021
LOCUS AK050021 3327 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus adult male liver tumor cDNA, RIKEN full-length enriched library, clone:C730003D12 product:DIPEPTIDYL PEPTIDASE 9 homolog [Homo sapiens], full insert sequence.
ACCESSION AK050021
VERSION AK050021.1 GI:26340743
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 3327)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.

```
FEATURES             Location/Qualifiers
     source          1. .3327
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:C730003D12"
                     /db_xref="taxon:10090"
                     /clone="C730003D12"
                     /sex="male"
                     /cell_type="tumor"
                     /tissue_type="liver"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             176. .2764
```

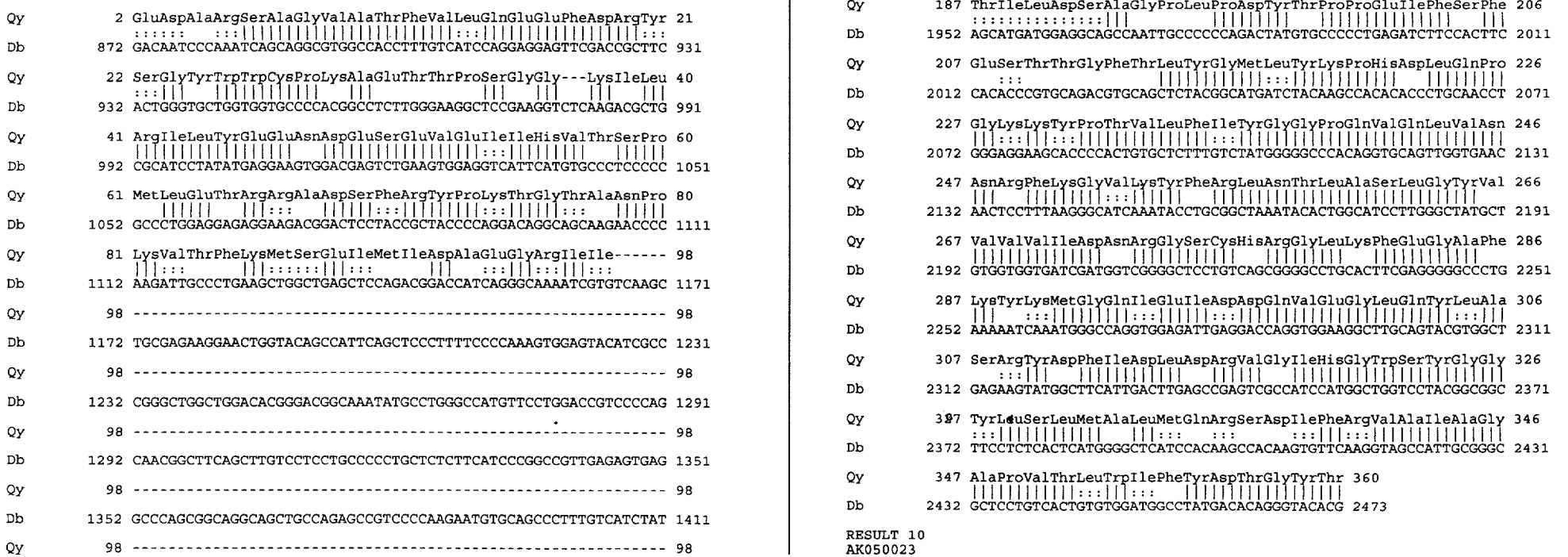
/note="unnamed protein product; DIPEPTIDYL PEPTIDASE 9 homolog [Homo sapiens] (SPTR|AAL47179, evidence: FASTY, 92%ID, 99.6%length, match=2580)
putative"
/codon_start=1
/protein_id="BAC34034.1"
/db_xref="GI:26340744"
/translation="MCSGVSPVEQVAAGDMDDTAARFCVQKHSWDGLRSIIHGSRKSS
GLIVSKAPHDFQFVQKPDESGPHSHRLYYLGMPYGSRENSLLYSEIPKKVRKEALLLL
SWKQMLDHFQATPHHGVYSREEELLRERKRLGVFGITSYDFHSESGLFLFQASNSLFH
CRDGGKNGFMVSPMKPLEIKTQCSGPRMDPKICPADPAFFSFINNSDLWVANIETGEE
RRLTFCHQGSAGVLDNPKSAGVATFVIQEEFDRFTGCWWCPTASWEGSEGLKTLRILY
EEVDESEVEVIHVPSPALEERKTDSYRYPRTGSKNPKIALKLAELQTDHQGKIVSSCE
KELVQPFSSLFPKVEYIARAGWTRDGKYAWAMFLDRPQQRLQLVLLPPALFIPAVESE
AQRQAAARAVPKNVQPFVIYEEVTNVWINVHDIFHPFPQAEGQQDFCFLRANECKTGF
CHLYRVTVELKTKDYDWTEPLSPTEDEFKCPIKEEVALTSGEWEVLSRHGSKIWVNEQ
TKLVYFQGTKDTPLEHHLYVVSYESAGEIVRLTTLGFSHSCSMSQSFDMFVSHYSSVS
TPPCVHVYKLSGPDDDPLHKQPRFWASMMEAANCPPDYVPPEIFHPHTRADVQLYGMI
YKPHTLQPGRKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNTLASLGYAVVVIDGRGSC
QRGLHFEGALKNQMGQVEIEDQVEGLQYVAEKYGFIDLSRVAIHGWSYGGFLSLMGLI
HKPQVFKVAIAGAPVTVWMAYDTGYTERYMDVPENNQQGYEAGSVALHVEKLPNEPNR
LLILHGFLDENVHFFHTNFLVSQLIRAGKPYQLQIYPNERHSIRCRESGEHYEVTLLH
FLQEHL"

BASE COUNT 702 a 1015 c 919 g 691 t
ORIGIN

Alignment Scores:

Pred. No.:	5.44e-127	Length:	3327
Score:	1204.50	Matches:	234
Percent Similarity:	53.75%	Conservative:	53
Best Local Similarity:	43.82%	Mismatches:	72
Query Match:	62.31%	Indels:	175
DB:	11	Gaps:	2

US-10-070-464-7 (1-360) x AK050021 (1-3327)



RESULT 10
AK050023

```
LOCUS       AK050023                3376 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus adult male liver tumor cDNA, RIKEN full-length
            enriched library, clone:C730003H19 product:DIPEPTIDYL PEPTIDASE 9
            homolog [Homo sapiens], full insert sequence.
ACCESSION   AK050023
VERSION     AK050023.1  GI:26093756
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3376)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissue was provided by William A. Held, Roswell Park Cancer
            Institute, Department of Molecular and Cellular Biology, Elm and
            Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully
            acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .3376
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:C730003H19"
                     /db_xref="taxon:10090"
                     /clone="C730003H19"
                     /sex="male"
                     /cell_type="tumor"
                     /tissue_type="liver"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     misc_feature    235. .2824
                     /note="DIPEPTIDYL PEPTIDASE 9 homolog [Homo sapiens]
                     (SPTR|AAL47179, evidence: FASTY, 92%ID, 99.6%length,
                     match=2580)
                     putative"
BASE COUNT      698 a   1046 c    932 g    700 t
ORIGIN
```

Alignment Scores:

Pred. No.:	5.56e-127	Length:	3376
Score:	1204.50	Matches:	234
Percent Similarity:	53.75%	Conservative:	53
Best Local Similarity:	43.82%	Mismatches:	72
Query Match:	62.31%	Indels:	175
DB:	11	Gaps:	2

US-10-070-464-7 (1-360) x AK050023 (1-3376)

```
Qy          2 GluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArgTyr 21
              ::::::   :::||||||||||||||||||||||||:::||||||||||||||||||:::
Db        932 GACAATCCCAAATCAGCAGGCGTGGCCACCTTTGTCATCCAGGAGGAGTTCGACCGCTTC 991

Qy         22 SerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly---LysIleLeu 40
              :::|||   ||||||||||||   |||           |||   |||   |||   |||
Db        992 ACTGGGTGCTGGTGGTGCCCCACGGCCTCTTGGGAAGGCTCCGAAGGTCTCAAGACGCTG 1051

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||   ||||||||||||||||||:::|||||||||   ||||||
Db       1052 CGCATCCTATATGAGGAAGTGGACGAGTCTGAAGTGGAGGTCATTCATGTGCCCTCCCCC 1111

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
                ||||||   |||:::   ||||||:::||||||||:::||||||:::   ||||||
```

```
Db       1112 GCCCTGGAGGAGAGGAAGACGGACTCCTACCGCTACCCCAGGACAGGCAGCAAGAACCCC 1171

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98

Db       1172 AAGATTGCCCTGAAGCTGGCTGAGCTCCAGACGGACCATCAGGGCAAAATCGTGTCAAGC 1231

Qy         98 ---------------------------------------------------------- 98

Db       1232 TGCGAGAAGGAACTGGTACAGCCATTCAGCTCCCTTTTCCCCAAAGTGGAGTACATCGCC 1291

Qy         98 ---------------------------------------------------------- 98

Db       1292 CGGGCTGGCTGGACACGGGACGGCAAATATGCCTGGGCCATGTTCCTGGACCGTCCCCAG 1351

Qy         98 ---------------------------------------------------------- 98

Db       1352 CAACGGCTTCAGCTTGTCCTCCTGCCCCCTGCTCTCTTCATCCCGGCCGTTGAGAGTGAG 1411

Qy         98 ---------------------------------------------------------- 98

Db       1412 GCCCAGCGGCAGGCAGCTGCCAGAGCCGTCCCCAAGAATGTGCAGCCCTTTGTCATCTAT 1471

Qy         98 ---------------------------------------------------------- 98

Db       1472 GAAGAAGTCACCAATGTCTGGATCAACGTCCACGACATCTTCCACCCGTTTCCTCAGGCT 1531

Qy         98 ---------------------------------------------------------- 98

Db       1532 GAGGGCCAGCAGGACTTTTGTTTCCTTCGTGCCAACGAATGCAAGACTGGCTTCTGCCAC 1591

Qy         98 ---------------------------------------------------------- 98

Db       1592 CTGTACAGGGTCACAGTGGAACTTAAAACCAAGGACTATGACTGGACGGAACCCCTCAGC 1651

Qy         98 ---------------------------------------------------------- 98

Db       1652 CCTACAGAAGATGAGTTTAAGTGCCCCATCAAGGAGGAGGTCGCCCTGACCAGTGGCGAG 1711

Qy         99 -----------------------------------ValAspGluValArgArgLeuVal 106

Db       1712 TGGGAGGTCTTGTCGAGGCATGGCTCCAAGATCTGGGTCAACGAGCAGACGAAGCTGGTG 1771

Qy        107 TyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrVal 126

Db       1772 TACTTTCAAGGTACAAAGGACACACCGCTGGAACATCACCTCTATGTGGTCAGCTACGAG 1831

Qy        127 AsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSer 146

Db       1832 TCAGCAGGCGAGATCGTGCGGCTCACCACGCTCGGCTTCTCCCACAGCTGCTCCATGAGC 1891

Qy        147 GlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSer 166

Db       1892 CAGAGCTTCGACATGTTCGTGAGTCACTACAGCAGTGTGAGCACGCCACCCTGTGTACAT 1951

Qy        167 LeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAla 186

Db       1952 GTGTACAAGCTGAGCGGCCCCGATGATGACCCACTGCACAAGCAACCACGCTTCTGGGCC 2011

Qy        187 ThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPhe 206

Db       2012 AGCATGATGGAGGCAGCCAATTGCCCCCCAGACTATGTGCCCCCTGAGATCTTCCACTTC 2071

Qy        207 GluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnPro 226

Db       2072 CACACCCGTGCAGACGTGCAGCTCTACGGCATGATCTACAAGCCACACACCCTGCAACCT 2131

Qy        227 GlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsn 246

Db       2132 GGGAGGAAGCACCCCACTGTGCTCTTTGTCTATGGGGGCCCACAGGTGCAGTTGGTGAAC 2191

Qy        247 AsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrVal 266

Db       2192 AACTCCTTTAAGGGCATCAAATACCTGCGGCTAAATACACTGGCATCCTTGGGCTATGCT 2251

Qy        267 ValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPhe 286

Db       2252 GTGGTGGTGATCGATGGTCGGGGCTCCTGTCAGCGGGGCCTGCACTTCGAGGGGGCCCTG 2311

Qy        287 LysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAla 306

Db       2312 AAAAATCAAATGGGCCAGGTGGAGATTGAGGACCAGGTGGAAGGCTTGCAGTACGTGGCT 2371

Qy        307 SerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGly 326

Db       2372 GAGAAGTATGGCTTCATTGACTTGAGCCGAGTCGCCATCCATGGCTGGTCCTACGGCGGC 2431

Qy        327 TyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGly 346

Db       2432 TTCCTCTCACTCATGGGGCTCATCCACAAGCCACAAGTGTTCAAGGTAGCCATTGCGGGC 2491

Qy        347 AlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360

Db       2492 GCTCCTGTCACTGTGTGGATGGCCTATGACACAGGGTACACG 2533
```

RESULT 11
AK078301

```
LOCUS       AK078301               3457 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus adult male olfactory brain cDNA, RIKEN full-length
            enriched library, clone:6430584G11 product:DIPEPTIDYL PEPTIDASE 9
            homolog [Homo sapiens], full insert sequence.
ACCESSION   AK078301
VERSION     AK078301.1  GI:26347124
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
```

```
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
   MEDLINE  21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3457)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..3457
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:6430584G11"
                     /db_xref="taxon:10090"
                     /clone="6430584G11"
                     /sex="male"
                     /tissue_type="olfactory brain"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             182..2770
                     /note="unnamed protein product; DIPEPTIDYL PEPTIDASE 9
                     homolog [Homo sapiens] (SPTR|AAL47179, evidence: FASTY,
                     92%ID, 99.6%length, match=2580)
                     putative"
                     /codon_start=1
                     /protein_id="BAC37211.1"
                     /db_xref="GI:26347125"
                     /translation="MCSGVSPVEQVAAGDMDDTAARFCVQKHSWDGLRSIIHGSRKSS
                     GLIVSKAPHDFQFVQKPDESGPHSHRLYYLGMPYGSRENSLLYSEIPKKVRKEALLLL
                     SWKQMLDHFQATPHHGVYSREEELLRERKRLGVFGITSYDFHSESGLFLFQASNSLFH
                     CRDGGKNGFMVSPMKPLEIKTQCSGPRMDPKICPADPAFFSFINNSDLWVANIETGEE
                     RRLTFCHQGSAGVLDNPKSAGVATFVIQEEFDRFTGCWWCPTASWEGSEGLKTLRILY
                     EEVDESEVEVIHVPSPALEERKTDSYRYPRTGSKNPKIALKLAELQTDHQGKIVSSCE
                     KELVQPFSSLFPKVEYIARAGWTRDGKYAWAMFLYRPQQRLQLVLLPPALFIPAVESE
                     AQRQAAARAVPKNVQPFVIYEEVTNVWINVHDIFHPFPQAEGQQDFCFLRANECKTGF
                     CHLYRVTVELKTKDYDWTEPLSPTEDEFKCPIKEEVALTSGEWEVLSRHGSKIWVNEQ
                     TKLVYFQGTKDTPLEHHLYVVSYESAGEIVRLTTLGFSHSCSMSQSFDMFVSHYSSVS
                     TPPCVHVYKLSGPDDDPLHKQPRFWASMMEAANCPPDYVPPEIFHFHTRADVQLYGMI
                     YKPHTLQPGRKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNTLASLGYAVVVIDGRGSC
                     QRGLHFEGALKNQMGQVEIEDQVEGLQYVAEKYGFIDLSRVAIHGWSYGGFLSLMGLI
                     HKPQVFKVAIAGAPVTVWMAYDTGYTERYMDVPENNKQGYEAGSVALHVEKLPNEPNR
                     LLILHGFLDENVHFFHTNFLVSQLIRAGKPYQLQIYPNERHSIRCRESGEHYEVTLLH
                     FLQEHL"
BASE COUNT      766 a   1033 c    926 g    732 t
ORIGIN

Alignment Scores:
Pred. No.:              5.77e-127       Length:         3457
Score:                  1204.50         Matches:        234
Percent Similarity:     53.75%          Conservative:   53
Best Local Similarity:  43.82%          Mismatches:     72
Query Match:            62.31%          Indels:         175
DB:                     11              Gaps:           2

US-10-070-464-7 (1-360) x AK078301 (1-3457)

Qy          2 GluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArgTyr 21
              :::::: :::||||||||||||||||||||||||:::|||||||||||||||||::: 
Db        878 GACAATCCCAAATCAGCAGGCGTGGCCACCTTTGTCATCCAGGAGGAGTTCGACCGCTTC 937

Qy         22 SerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly---LysIleLeu 40
              :::||| ||||||||||  |||            |||  |||  |||  |||
Db        938 ACTGGGTGCTGGTGGTGCCCCACGGCCTCTTGGGAAGGCTCCGAAGGTCTCAAGACGCTG 997

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||| ||||||||||||||||:::|||||||| |||||
Db        998 CGCATCCTATATGAGGAAGTGGACGAGTCTGAAGTGGAGGTCATTCATGTGCCCTCCCCC 1057

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
               |||||| |||:::  ||||||:::||||||||:::||||||:::  ||||||
Db       1058 GCCCTGGAGGAGAGGAAGACGGACTCCTACCGCTACCCCAGGACAGGCAGCAAGAACCCC 1117

Qy         81 LysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIleIle------ 98
              |||:::   |||::::::|||:::     |||   :::|||:::|||:::
Db       1118 AAGATTGCCCTGAAGCTGGCTGAGCTCCAGACGGACCATCAGGGCAAAATCGTGTCAAGC 1177

Qy         98 ------------------------------------------------------------ 98

Db       1178 TGCGAGAAGGAACTGGTACAGCCATTCAGCTCCCTTTTCCCCAAAGTGGAGTACATCGCC 1237

Qy         98 ------------------------------------------------------------ 98

Db       1238 CGGGCTGGCTGGACACGGGACGGCAAATATGCCTGGGCCATGTTCCTGTACCGTCCCCAG 1297

Qy         98 ------------------------------------------------------------ 98

Db       1298 CAACGGCTTCAGCTTGTCCTCCTACCCCCTGCTCTCTTCATCCCGGCCGTTGAGAGTGAG 1357

Qy         98 ------------------------------------------------------------ 98

Db       1358 GCCCAGCGGCAGGCAGCTGCCAGAGCCGTCCCCAAGAATGTGCAGCCCTTTGTCATCTAT 1417

Qy         98 ------------------------------------------------------------ 98

Db       1418 GAAGAAGTCACCAATGTCTGGATCAACGTCCACGACATCTTCCACCCGTTTCCTCAGGCT 1477

Qy         98 ------------------------------------------------------------ 98

Db       1478 GAGGGCCAGCAGGACTTTTGTTTCCTTCGTGCCAACGAATGCAAGACTGGCTTCTGCCAC 1537

Qy         98 ------------------------------------------------------------ 98

Db       1538 CTGTACAGGGTCACAGTGGAACTTAAAACCAAGGACTATGACTGGACGGAACCCCTCAGC 1597

Qy         98 ------------------------------------------------------------ 98
```

```
Db       1598 CCTACAGAAGATGAGTTTAAGTGCCCCATCAAGGAGGAGGTCGCCCTGACCAGTGGCGAG 1657

Qy         99 ------------------------------------ValAspGluValArgArgLeuVal 106
                                                    |||:::|||      :::||||||
Db       1658 TGGGAGGTCTTGTCGAGGCATGGCTCCAAGATCTGGGTCAACGAGCAGACGAAGCTGGTG 1717

Qy        107 TyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSerTyrVal 126
              |||||||:::|||||||||||||:::|||||||||||||||||||||||||||||||||
Db       1718 TACTTTCAAGGTACAAAGGACACACCGCTGGAACATCACCTCTATGTGGTCAGCTACGAG 1777

Qy        127 AsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCysIleSer 146
              :::   ||||||:::   |||||||||      |||:::||||||||||||   :::|||
Db       1778 TCAGCAGGCGAGATCGTGCGGCTCACCACGCTCGGCTTCTCCCACAGCTGCTCCATGAGC 1837

Qy        147 GlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCysValSer 166
              |||     |||   |||:::|||   ||||||:::         |||   ||||||
Db       1838 CAGAGCTTCGACATGTTCGTGAGTCACTACAGCAGTGTGAGCACGCCACCCTGTGTACAT 1897

Qy        167 LeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPheTrpAla 186
              :::||||||||||||   |||:::|||||||||      |||        |||||||||
Db       1898 GTGTACAAGCTGAGCGGCCCCGATGATGACCCACTGCACAAGCAACCACGCTTCTGGGCC 1957

Qy        187 ThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePheSerPhe 206
              :::::::::::::::|||         |||||||||   ||||||||||||||   |||
Db       1958 AGCATGATGGAGGCAGCCAATTGCCCCCCAGACTATGTGCCCCCTGAGATCTTCCACTTC 2017

Qy        207 GluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeuGlnPro 226
              :::                  ||||||||||||:::||||||||||||   |||||||||
Db       2018 CACACCCGTGCAGACGTGCAGCTCTACGGCATGATCTACAAGCCACACACCCTGCAACCT 2077

Qy        227 GlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeuValAsn 246
              |||:::|||:::|||||||||||||||:::|||||||||||||||||||||||||||||
Db       2078 GGGAGGAAGCACCCCACTGTGCTCTTTGTCTATGGGGGCCCACAGGTGCAGTTGGTGAAC 2137

Qy        247 AsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGlyTyrVal 266
              |||   |||||||||:::||||||   ||||||||||||||||||||||||||||||||
Db       2138 AACTCCTTTAAGGGCATCAAATACCTGCGGCTAAATACACTGGCATCCTTGGGCTATGCT 2197

Qy        267 ValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGlyAlaPhe 286
              ||||||||||||||   ||||||||||||   |||||||||   ||||||||||||||
Db       2198 GTGGTGGTGATCGATGGTCGGGGCTCCTGTCAGCGGGGCCTGCACTTCGAGGGGGCCCTG 2257

Qy        287 LysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyrLeuAla 306
              |||   :::|||||||||:::||||||:::||||||||||||||||||||:::|||
Db       2258 AAAAATCAAATGGGCCAGGTGGAGATTGAGGACCAGGTGGAAGGCTTGCAGTACGTGGCT 2317

Qy        307 SerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyrGlyGly 326
                 :::|||   ||||||||||||   ||||||   ||||||||||||||||||||||||
Db       2318 GAGAAGTATGGCTTCATTGACTTGAGCCGAGTCGCCATCCATGGCTGGTCCTACGGCGGC 2377

Qy        327 TyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGly 346
              :::||||||||||||   |||:::   :::      :::|||:::||||||||||||||||
Db       2378 TTCCTCTCACTCATGGGGCTCATCCACAAGCCACAAGTGTTCAAGGTAGCCATTGCGGGC 2437

Qy        347 AlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThr 360
              ||||||||||||:::|||:::   ||||||||||||||||||
Db       2438 GCTCCTGTCACTGTGTGGATGGCCTATGACACAGGGTACACG 2479

RESULT 12
BU239476
LOCUS       BU239476               1082 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  603322338F1 CSEQCHN33 Gallus gallus cDNA clone ChEST24919 5', mRNA
            sequence.
ACCESSION   BU239476
VERSION     BU239476.1  GI:25485538
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1082)
  AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
  TITLE     A Comprehensive Collection of Chicken cDNAs
  JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
  MEDLINE   22335534
   PUBMED   12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology (UMIST
            )
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .1082
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="Compton Line 15I"
                     /db_xref="taxon:9031"
                     /clone="ChEST24919"
                     /sex="Female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN33"
                     /note="Organ: liver; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: Not1; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with EcoRI
                     , size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."
BASE COUNT      289 a    240 c    280 g    273 t
ORIGIN

Alignment Scores:
Pred. No.: 0                 3.05e-118       Length:          1082
Score:                       1122.00         Matches:         217
Percent Similarity:          88.01%          Conservative:    18
Best Local Similarity:       81.27%          Mismatches:      27
Query Match:                 58.04%          Indels:          5
DB:                          13              Gaps:            2

US-10-070-464-7 (1-360) x BU239476 (1-1082)

Qy         97 IleIleValAspGluValArgArgLeuValTyrPheGluGlyThrLysAspSerProLeu 116
              |||   |||||||||   ::::::||||||||||||:::|||||||||||||||||||||
Db        107 ATCTATGTAGATGAAGCCAAAAAACTGGTCTATTTTCAAGGCACGAAAGACTCTCCTTTA 166

Qy        117 GluHisHisLeuTyrValValSerTyrValAsnProGlyGluValThrArgLeuThrAsp 136
              ||||||||||||||||||||||:::|||   |||||||||||||||   |||||||:::
Db        167 GAACATCACTTGTACGTTGTTAACTACAAGAATCCTGGAGAAGTAAAGCGACTGACAGAA 226

Qy        137 ArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyr 156
              ||||||||||||||:::||||||:::||||||   ||||||   |||||||||||||||
Db        227 CGTGGTTACTCTCATGCCTGCTGTGTCAGCCAGGATTGTGACATGTTCATCAGCAAGTAC 286

Qy        157 SerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAsp 176
              ||||||||||||||||||||||||||||||||||:::|||:::      |||||||||
Db        287 AGCAATCAGAAGAACCCACACTGTGTGTCCCTTTACCGGCTGACAGGACATGAAGATGAT 346

Qy        177 ProThrCysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuPro 196
                    :::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        347 GCAGCTCAAAGGACAAAGGAATTCTGGGCTACAATTTTGGATTCAGCAGGCCCTCTTCCT 406

Qy        197 AspTyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGly 216
              ||||||   |||||||||:::||||||||||||||:::||||||||||||||||||||||
```

```
Db        407 GATTACATTCCCCCAGAAGTGTTCTCCTTTGAGAGCTCCACGGGGTTTACACTCTATGGG 466

Qy        217 MetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIle 236
              |||:::|||||||||||:::|||||||||||||||||||||||||||||:::||||||
Db        467 ATGATGTACAAACCTCACAATCTGCAACCTGGAAAGAAGTACCCTACTGTGATCTTCATC 526

Qy        237 TyrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArg 256
              ||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        527 TATGGAGGCCCTCAGGTGCAGCTAGTGAACAATCGATTTAAAGGACTCAAATATTTCCGA 586

Qy        257 LeuAsnThrLeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCys 276
              |||||||||||||||||||||||||||||||   |||||||||||||||||||||||||
Db        587 TTGAACACCTTGGCCTCTTTAGGCTACGTTGTTGCTGTTATTGACAACCGGGGGTCCTGC 646

Qy        277 HisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAsp 296
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        647 CACCGAGGGCTGAAGTTTGAAGGAGCCTTTAAATACAAAATGGGACAAATAGAAATTGAT 706

Qy        297 AspGlnValGlu-GlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspAr 316
              |||||||||||| ||||||   |||          :::|||||||||||||||      ||
Db        707 GACCAAGTGGAAGGGTCTGCACTACTTTGGCGTCCAGTATGACTTCATTGATTGGCTTCG 766

Qy        316 gValGlyIleHisGlyTrpSerTyrGlyGlyTyr---LeuSerLeuMetAlaLeuMetGl 335
              |||||||||||||||||||||||||||||||     ||||||   ||||||
Db        767 GGTTGGCATTCATGGGTGGTCCTATGGAGGGCTAACTCTCTCTTATGGCGTTTTAATGCC 826

Qy        335 nArgSer-AspIlePheArgValAlaIleAlaGlyAla-ProValThrLeuTrp---Ile 353
              ||||||  |||||||||||||||||||||||||||::: ||||||||||||      |||
Db        827 GAGGTCAGGATATCTTCAGGGTTGCCATCGCCGGATCACCCGGTAACGCTGGGGGAGATT 886

Qy        354 PheTyrAspThrGly 358
              |||||||||
Db        887 CCTATAGACACGGGT 901
```

```
RESULT 13
BE888665
LOCUS       BE888665                674 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601513061F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3914582 5',
            mRNA sequence.
ACCESSION   BE888665
VERSION     BE888665.1  GI:10345197
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 674)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9736  row: i  column: 15
            High quality sequence stop: 662.
FEATURES             Location/Qualifiers
     source          1. .674
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3914582"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_71"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2.1 kb.  "
```

```
BASE COUNT      185 a    143 c    148 g    198 t
ORIGIN

Alignment Scores:
Pred. No.:              5.52e-118       Length:          674
Score:                  1117.00         Matches:         219
Percent Similarity:     96.90%          Conservative:    0
Best Local Similarity:  96.90%          Mismatches:      6
Query Match:            57.79%          Indels:          4
DB:                     10              Gaps:            0

US-10-070-464-7 (1-360) x BE888665 (1-674)

Qy        115 ProLeuGluHisHisLeuTyrValValSerTyrValAsnProGlyGluValThrArgLeu 134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 CCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAATCCTGGAGAGGTGACAAGGCTG 61

Qy        135 ThrAspArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSer 154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 ACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAGCACTGTGACTTCTTTATAAGT 121

Qy        155 LysTyrSerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSerProGlu 174
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTTTACAAGCTATCAAGTCCTGAA 181

Qy        175 AspAspProThrCysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyPro 194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACCATTTTGGATTCAGCAGGTCCT 241

Qy        195 LeuProAspTyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeu 214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 CTTCCTGACTATACTCCTCCAGAAATTTTCTCTTTTGAAAGTACTACTGGATTTACATTG 301

Qy        215 TyrGlyMetLeuTyrLysProHisAspLeuGlnProGlyLysLysTyrProThrValLeu 234
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 TATGGGATGCTCTACAAGCCTCATGATCTACAGCCTGGAAAGAAATATCCTACTGTGCTG 361

Qy        235 PheIleTyrGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyr 254
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 TTCATATATGGTGGTCCTCAGGTGCAGTTGGTGAATAATCGGTTTAAAGGAGTCAAGTAT 421

Qy        255 PheArgLeuAsnThrLeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGly 274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 TTCCGCTTGAATACCCTAGCCTCTCTAGGTTATGTGGTTGTAGTGATAGACAACAGGGGA 481

Qy        275 SerCysHisArgGlyLeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGlu 294
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 TCCTGTCACCGAGGGCTTAAATTTGAAGGCGCCTTTAAATATAAATGGGTCAACATAGAA 541

Qy        295 IleAspAspGlnValGluGly-LeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLe 314
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        542 ATTGACGATCAGGTGGAAGGAACTCCAATATCTAGCTTCTCGATATGATTTCATTGACTT 601

Qy        314 uAspArgValGlyIleHisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMe 334
              ||||||||||||||||||   ||||||||||||||||||   ||||||||||||||||||
Db        602 AGATCGTGTGGGCATCCACGG-TGGTCCTATGGAGGATACTC-TCCCTGATGGCATTAAT 659

Qy        334 tGlnArgSerAspIle 339
              ||||   |||||||||
Db        660 GCAAAG-TCAGATATC 674
```

```
RESULT 14
BM910838
LOCUS       BM910838                1090 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6615934 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454402
            5', mRNA sequence.
ACCESSION   BM910838
VERSION     BM910838.1  GI:19361217
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 1090)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1949  row: h  column: 19
            High quality sequence stop: 622.
FEATURES             Location/Qualifiers
     source          1..1090
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5454402"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_98"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      317 a    254 c    240 g    278 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              1.11e-116      Length:        1090
Score:                  1108.50        Matches:       237
Percent Similarity:     77.92%         Conservative:  10
Best Local Similarity:  74.76%         Mismatches:    54
Query Match:            57.35%         Indels:        16
DB:                     12             Gaps:          4

US-10-070-464-7 (1-360) x BM910838 (1-1090)

Qy          1 GluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGluPheAspArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         82 GAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAATTTGATAGA 141

Qy         21 TyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGlyLysIleLeu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        142 TATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGTAAAATTCTT 201

Qy         41 ArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisValThrSerPro 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        202 AGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTTACATCCCCT 261

Qy         61 MetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThrAlaAsnPro 80
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        262 ATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTGATTTCAAGTGT 321

Qy         81 LysValThrPheLysMetSer------------------GluIleMetIleAspAlaGlu 94
                :::      ::::::::                   |||:::::
Db        322 CCTATCAAAGAGGAGATAGCAATTACCAGTGGTGAATGGGAAGTTCTTGGCCGGCATGGA 381

Qy         95 GlyArgIleIleValAspGluValArgArgLeuValTyrPheGluGlyThrLysAspSer 114
                 |||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db        382 TCTAATATCCAAGTTGATGAAGTCAGAAGGCTGGTATATTTTGAAGGCACCAAAGACTCC 441

Qy        115 ProLeuGluHisHisLeuTyrValValSerTyrValAsnProGlyGluValThrArgLeu 134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        442 CCTTTAGAGCATCACCTGTACGTAGTCAGTTACGTAAATCCTGGAGAGGTGACAAGGCTG 501

Qy        135 ThrAspArgGlyTyrSerHisSerCysCysIleSerGlnHisCysAspPhePheIleSer 154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        502 ACTGACCGTGGCTACTCACATTCTTGCTGCATCAGTCAGCACTGTGACTTCTTTATAAGT 561

Qy        155 LysTyrSerAsnGlnLysAsnProHisCysValSerLeuTyrLysLeuSerSer-ProGl 174
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        562 AAGTATAGTAACCAGAAGAATCCACACTGTGTGTCCCTTTACAAGCTATCAAGTCCCTGA 621

Qy        174 uAspAspProThrCysLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGly-P 194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        622 AGATGACCCAACTTGCAAAACAAAGGAATTTTGGGCCACCATTTTGGATTCAGCAGGNCC 681

Qy        194 roLeuProAspTyrThrProProGluIle-PheSerPheGluSerThrThrGly-PheTh 213
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||| |||||
Db        682 CTCTTCCTGACTATACTCCTCCAGAAATTTTTCTCTTTTGAAAGTACTACCGGAATTTAC 741

Qy        213 rLeuTyrGlyMetLeu-TyrLysProHisAspLeuGlnProGlyLysLysTyrProThrV 233
              ||||    |||||||| |||:::|||||||||||    ||||||||||||||||||||
Db        742 ATTGAATGGGATGCTCCTACCAGCCTCATGATCTTACGCCTGGAAAGAAATATCCTACTG 801

Qy        233 alLeuPhe-IleTyrGlyGlyProGlnValGlnLeuVal-AsnAsnArgPheLysGlyVa 252
                |||||     |||   ||||||||         ||||||          ||||||    ||||||::
Db        802 TGCTTGTCCATAATGGGGGGTCCTCCGGGGCAGTTGGCGGAAAAACCGGCTTAAAGGAAT 861

Qy        252 lLysTyrPheArgLeuAsnThrLeuAlaSerLeu---GlyTyrValValValValIleAs 271
               :||||||  |||           |||         ||:::       |||       |||
Db        862 CAAGTATTTCCCCCTTGAAAACCCTCAACCCCTCTAAGGTTTGGGGTGGCACGCATAAC 921

Qy        271 pAsnArgGly---SerCysHisArgGly---LeuLysPheGluGlyAlaPheLysTyrLy 289
                           |||   |||          :::   |||   |||||
Db        922 ACAAGAGGCGGTCTCTGCCCACCCAGCCGTTATACATTTAGAGGGGCCTTGTAAATAAA 981

Qy        289 sMetGlyGlnIleGluIleAspAspGlnValGluGly 301
                  |||               |||        |||
Db        982 CAAAGGCGCCGAACGACAACAGGACTACTCGGCGGGT 1018

RESULT 15
BG165205
LOCUS       BG165205                 639 bp    mRNA    linear   EST 06-FEB-2001
DEFINITION  602344074F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4453987 5',
            mRNA sequence.
ACCESSION   BG165205
VERSION     BG165205.1  GI:12671908
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 639)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10245  row: d  column: 20
            High quality sequence stop: 639.
FEATURES             Location/Qualifiers
     source          1..639
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4453987"
                     /tissue_type="hypernephroma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_89"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
```

Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 169 a 137 c 142 g 191 t
ORIGIN

Alignment Scores:

Pred. No.:	6.03e-116	Length:	639
Score:	1099.00	Matches:	210
Percent Similarity:	98.59%	Conservative:	0
Best Local Similarity:	98.59%	Mismatches:	2
Query Match:	56.85%	Indels:	2
DB:	10	Gaps:	0

US-10-070-464-7 (1-360) x BG165205 (1-639)

```
Qy        140 SerHisSerCysCysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGln 159
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 TCACATTCTTGCTGCATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAG 61

Qy        160 LysAsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCys 179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 AAGAATCCACACTGTGTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGC 121

Qy        180 LysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThr 199
              |||||||||||||||||||||    ||||||||||||||||||||||||||||||||||||||
Db        122 AAAACAAAGGAATTTTGGGCCACCAT-TTGGATTCAGCAGGTCCTCTTCCTGACTATACT 180

Qy        200 ProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyr 219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCTCCAGAAATTTTCTCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTAC 240

Qy        220 LysProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAGCCTCATGATCTACAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGT 300

Qy        240 ProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThr 259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCTCAGGTGCAGTTGGTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACC 360

Qy        260 LeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerCysHisArgGly 279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTAGCCTCTCTAGGTTATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGG 420

Qy        280 LeuLysPheGluGlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnVal 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTAAATTTGAAGGCGCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTG 480

Qy        300 GluGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIle 319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAAGGACTCCAATATCTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATC 540

Qy        320 HisGlyTrpSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIle 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CACGGCTGGTCCTATGGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATC 600

Qy        340 Phe-ArgValAlaIleAlaGlyAlaProValThrLeu 351
              |||   ||||||||||||||||||||||||||||||
Db        601 TTCAGCGGTTGCTATTGCTGGGGCCCCAGTCACTCTG 637
```

Search completed: October 16, 2003, 03:25:54
Job time : 2181.86 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 15, 2003, 15:44:49 ; Search time 53.7858 Seconds
(without alignments)
2602.858 Million cell updates/sec

Title: US-10-070-464-1
Perfect score: 4700
Sequence: 1 MAAAMETEQLGVEIFETADC.....HLLHYLQENLGSRIAALKVI 882

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4700	100.0	882	22	AAB47187	Human DPP8. Homo
2	4700	100.0	882	23	AAE24170	Human dipeptidyl p
3	4700	100.0	882	23	ABG61591	Human DPPIV relate
4	4700	100.0	882	23	AAU74749	Human protease PRT
5	4700	100.0	882	23	AAG78415	Amino acid sequenc
6	4700	100.0	882	24	ABU07720	Human serine prote
7	4118	87.6	782	23	ABB97361	Novel human protei
8	3771	80.2	724	23	ABB97362	Novel human protei
9	3504	74.6	658	23	ABG61600	Human DPRP-1 splic
10	3504	74.6	661	23	ABG61596	Human DPRP-1 splic
11	3504	74.6	690	23	ABG61594	Human DPRP-1 splic
12	3333.5	70.9	632	22	AAB93565	Human protein sequ
13	3236	68.9	613	23	ABG61601	Human DPRP-1 splic
14	2870	61.1	863	23	ABG61592	Human DPPIV relate
15	2870	61.1	892	23	ABB98134	Human PMMM Incyte
16	2870	61.1	892	23	ABG61602	Human DPRP-2 splic
17	2870	61.1	892	23	ABG61604	Human DPRP-2 splic
18	2863	60.9	969	23	AAE24168	Human dipeptidyl p
19	2835	60.3	830	23	AAE24171	Human dipeptidyl p
20	2833	60.3	869	23	AAE24169	Alternative versio
21	2820.5	60.0	879	23	ABG61607	Human DPRP-2 splic
22	2820.5	60.0	879	23	ABG61608	Human DPRP-2 splic
23	2713.5	57.7	847	23	AAE23875	Murine dipeptidyl
24	2547.5	54.2	580	23	AAE14337	Human protease PRT
25	2422	51.5	465	22	AAB47189	Human DPP8 318Thr-
26	2413.5	51.4	737	22	AAM38724	Human polypeptide
27	2406	51.2	832	23	ABG61605	Human DPRP-2 splic
28	2406	51.2	832	23	ABG61606	Human DPRP-2 splic
29	2356.5	50.1	819	23	ABG61609	Human DPRP-2 splic
30	2356.5	50.1	819	23	ABG61610	Human DPRP-2 splic
31	2233.5	47.5	683	22	AAM40510	Human polypeptide
32	1914.5	40.7	720	21	AAB41626	Human ORFX ORF1390
33	1836.5	39.1	360	22	AAB47190	Human DPP8 244Glu-
34	1808	38.5	358	23	ABG61597	Human DPRP-1 splic
35	1754.5	37.3	497	23	ABG64845	Human albumin fusi
36	1754.5	37.3	497	23	AAU96169	Human secreted pro
37	1750.5	37.2	497	23	ABG64844	Human albumin fusi
38	1750.5	37.2	497	23	AAU96192	Human secreted pro
39	1645.5	35.0	310	22	AAB47188	Human DPP8 524Phe-
40	1645.5	35.0	310	23	ABB08994	Human dipeptidyl p
41	1643.5	35.0	508	21	AAB42928	Human ORFX ORF2692
42	1599.5	34.0	1042	22	ABB60137	Drosophila melanog
43	1599.5	34.0	1102	22	ABB62029	Drosophila melanog
44	1400	29.8	518	21	AAY90299	Human peptidase, H
45	1391	29.6	518	23	ABG61603	Human DPRP-2 splic

ALIGNMENTS

RESULT 1
AAB47187
ID AAB47187 standard; Protein; 882 AA.
XX
AC AAB47187;
XX
DT 29-JUN-2001 (first entry)
XX
DE Human DPP8.
XX
KW Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
KW dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;
KW growth hormone deficiency; glucose level; mucosal regeneration;
KW non-insulin dependent diabetes mellitus;glucose intolerance;
KW immunosuppression.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Active-site 739
FT /note= "Forms part of Ser-Asp-His catalytic triad"
FT Active-site 817
FT /note= "Forms part of Ser-Asp-His catalytic triad"
FT Active-site 849
FT /note= "Forms part of Ser-Asp-His catalytic triad"
XX
PN WO200119866-A1.
XX
PD 22-MAR-2001.
XX
PF 11-SEP-2000; 2000WO-AU01085.
XX

PR 10-SEP-1999; 99AU-0002762.
PR 18-FEB-2000; 2000AU-0005709.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Abbott CA, Gorell MD;
XX
DR WPI; 2001-281520/29.
DR N-PSDB; AAC85694.
XX
PT New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
PT substrates, identifying inhibitors of DPP8 catalytic activity which
PT have therapeutic uses, and for detecting activated T cells -
XX
PS Claim 1; Fig 2; 78pp; English.
XX
CC This sequence represents human dipeptidyl aminopeptidase (DPP8).
CC DPP8 has substrate specificity for H-Ala-Pro-pNA, H-Gly-Pro-pNA and
CC H-Arg-Pro-pNA. Therefore, it is a prolyl oligopeptidase and a
CC dipeptidyl peptidase, because it is capable of hydrolysing the
CC peptide bond C-terminal to Pro in each of these compounds. DPP8
CC is homologous with human DPPIV. DPP8 is useful for cleaving a
CC substrate, and for detecting an activated T cell which involves
CC measuring the level of DPP8 gene expression in a T cell. The level
CC of DPP8 expression is detected by detecting the amount of DPP8 RNA
CC in the cell. It is also useful for identifying a molecule capable
CC of inhibiting the cleavage of the substrate by DPP8. Molecules
CC identified as inhibiting DPP8 catalytic activity may be useful for
CC treating diarrhoea, growth hormone deficiency, lowering glucose levels
CC in non-insulin dependent diabetes mellitus and other disorders
CC involving glucose intolerance, enhancing mucosal regeneration and
CC as immunosuppressants.
XX
SQ Sequence 882 AA;

Query Match 100.0%; Score 4700; DB 22; Length 882;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660

Qy        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720

Qy        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780

Qy        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840

Qy        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
              |||||||||||||||||||||||||||||||||||||||||||
Db        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
```

RESULT 2
AAE24170
ID AAE24170 standard; Protein; 882 AA.
XX
AC AAE24170;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human dipeptidyl peptidase 8 (DPP8) protein.
XX
KW Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
KW autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;
KW graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
KW antiviral; enzyme.
XX
OS Homo sapiens.
XX
PN WO200234900-A1.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-AU01388.
XX
PR 27-OCT-2000; 2000AU-0001078.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Abbott CA, Gorrell MD;
XX
DR WPI; 2002-454646/48.
DR N-PSDB; AAD38956.
XX
PT New dipeptidyl peptidase (DPP) peptides, useful for screening
PT inhibitors of DPP catalytic activity, which may be employed to treat
PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PT rejection and HIV infection -
XX
PS Example; Fig 1; 91pp; English.
XX
CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
CC polynucleotides encoding such proteins. The DPP peptides are useful for
CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
CC rejection and HIV (human immuno deficiency virus) infection. The present
CC sequence is human DPP8 protein.
XX

```
SQ   Sequence   882 AA;

  Query Match             100.0%;  Score 4700;  DB 23;  Length 882;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  882;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWININHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWININHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660

Qy        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720

Qy        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780

Qy        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840

Qy        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
              |||||||||||||||||||||||||||||||||||||||||||
Db        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882


RESULT 3
ABG61591
ID   ABG61591 standard; Protein; 882 AA.
XX
AC   ABG61591;
XX
DT   12-AUG-2002  (first entry)
XX
DE   Human DPPIV related serine protease DPRP-1.
XX
KW   Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW   DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW   diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW   heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW   ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW   dyskinesia; reproductive disorder; inflammatory disorder;
KW   metabolic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200231134-A2.
XX
PD   18-APR-2002.
XX
PF   12-OCT-2001; 2001WO-US31874.
XX
PR   12-OCT-2000; 2000US-240117P.
XX
PA   (FERR ) FERRING BV.
XX
PI   Qi S,  Akinsanya KO,  Riviere PJ,  Junien J;
XX
DR   WPI; 2002-444178/47.
DR   N-PSDB; ABK83322.
XX
PT   New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT   the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT   viral infections, cancers, allergies, neurological disorders, or pain
PT   -
XX
PS   Claim 17; Fig 1; 113pp; English.
XX
CC   The present invention relates to the isolation of novel human serine
CC   proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC   proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC   and nucleic acids encoding them are useful for treating infections
CC   such as fungal, bacterial, protozoan and viral infections, particularly
CC   infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC   pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC   bulimia, Parkinson's disease, acute heart failure, hypotension,
CC   hypertension, urinary retention, osteoporosis, angina pectoris,
CC   stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC   psychotic and neurological disorders (e.g. anxiety, dementia, or
CC   schizophrenia), and dyskinesias. These may also be used in discovering
CC   therapeutic agents for the treatment of reproductive, inflammatory and
CC   metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ   Sequence   882 AA;

  Query Match             100.0%;  Score 4700;  DB 23;  Length 882;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  882;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
```

Good for Utility

```
Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660

Qy        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720

Qy        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780

Qy        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840

Qy        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
              |||||||||||||||||||||||||||||||||||||||||||
Db        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
```

RESULT 4
AAU74749
ID AAU74749 standard; Protein; 882 AA.
XX
AC AAU74749;
XX
DT 09-APR-2002 (first entry)
XX
DE Human protease PRTS-9 protein sequence.
XX
KW Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KW cell proliferative disorder; developmental disorder; epilepsy;
KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KW reproductive disorder; endometriosis.
XX
OS Homo sapiens.
XX
PN WO200198468-A2.
XX
PD 27-DEC-2001.
XX
PF 13-JUN-2001; 2001WO-US19178.
XX
PR 16-JUN-2000; 2000US-212336P.
PR 22-JUN-2000; 2000US-213955P.
PR 29-JUN-2000; 2000US-215396P.
PR 07-JUL-2000; 2000US-216821P.
PR 14-JUL-2000; 2000US-218946P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM;
PI Delegeane AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
PI Walia NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT;
PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
PI Kallick DA;
XX
DR WPI; 2002-090437/12.
DR N-PSDB; ABK12892.
XX
PT Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
PT proliferative (e.g. cancer)disorders -
XX
PS Claim 1; Page 140-142; 177pp; English.
XX
CC The present invention relates to twenty one new human proteases,
CC referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
CC polypeptides of the invention are useful in the diagnosis, treatment and
CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
CC myocardial infarction, autoimmune/inflammatory e.g. acquired
CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
CC endometriosis disorders. Numerous other examples of each disorder are
CC given in the specification. The present protein sequence represents
CC the human protease PRTS-9 protein of the invention.
XX
SQ Sequence 882 AA;

Query Match 100.0%; Score 4700; DB 23; Length 882;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
```

```
Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660

Qy        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720

Qy        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780

Qy        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840

Qy        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
              |||||||||||||||||||||||||||||||||||||||||||
Db        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
```

```
RESULT 5
AAG78415
ID   AAG78415 standard; Protein; 882 AA.
XX
AC   AAG78415;
XX
DT   12-APR-2002  (first entry)
XX
DE   Amino acid sequence of 21953 human prolyl oligopeptidase.
XX
KW   21953 prolyl oligopeptidase; antibody; proline; endopeptidase;
KW   cancer; cardiovascular disease; autoimmune disease; atopic allergy;
KW   neuronal disorder; vascular disorder; prostate disorder; cytostatic;
KW   antidiabetic; antiarthritic; antiasthmatic; antiinflammatory;
KW   diabetes mellitus; arthritis; multiple sclerosis; asthma;
KW   Grave's disease; neuronal disorder; demyelinating disease.
XX
OS   Homo sapiens.
XX
PN   WO200179473-A2.
XX
PD   25-OCT-2001.
XX
PF   11-APR-2001; 2001WO-US40483.
XX
PR   18-APR-2000; 2000US-197508P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Meyers RA,  Williamson M;
XX
DR   WPI; 2002-034353/04.
DR   N-PSDB; AAH99934.
XX
PT   New polypeptides 21953, member of human prolyl oligopeptidase family,
PT   useful as diagnostic targets and therapeutic agents for controlling
PT   cancer, lymphoma and leukemia  -
XX
PS   Claim 1; Page 102-103; 121pp; English.
XX
CC   This invention relates to an isolated 21953 human prolyl
CC   oligopeptidase. Which is cytostatic, antidiabetic, antiarthritic,
CC   neuroprotective, antithyroid, dermatological, antipsoriatic,
CC   antiasthmatic, ophthalmological, antiinflammatory, nootropic,
CC   antiparkinsonian, anticonvulsant, gynaecological, vasotropic,
CC   antianginal, cardiant, antiatherosclerotic, anorectic and
CC   metabolic in its action. Uses include gene therapy, expression or
CC   activity of 21953 protein modulator, it is useful for identifying a
CC   compound which binds to it and can be used in preventing, treating
CC   or detecting a cellular proliferative or differentiative disorder.
CC   The 21953 molecules can act as novel diagnostic targets and therapeutic
CC   agents for controlling disorders associated with the aberrant activity
CC   or degradation of peptide hormones e.g., disorders associated with cell
CC   differentiation and proliferation such as cancer, immune function,
CC   reproductive, neurological and cardiovascular function. The 21953
CC   molecules are thus useful for treating and preventing cellular
CC   proliferative and differentiative disorders, haematopoietic neoplastic
CC   disorders, immune disorders such as autoimmune diseases, diabetes
CC   mellitus, arthritis, multiple sclerosis, asthma, Grave's disease,
CC   neuronal disorders, demyelinating diseases, vascular disorders and
CC   metabolism or pain disorders. This sequence represents the amino
CC   acid sequence of 21953 human prolyl oligopeptidase.
XX
SQ   Sequence   882 AA;
```

```
  Query Match             100.0%;  Score 4700;  DB 23;  Length 882;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  882;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660

Qy        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720
```

```
Qy        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780

Qy        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840

Qy        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
              |||||||||||||||||||||||||||||||||||||||||||
Db        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882


RESULT 6
ABU07720
ID   ABU07720 standard; Protein; 882 AA.
XX
AC   ABU07720;
XX
DT   19-MAY-2003  (first entry)
XX
DE   Human serine protease HIPHUM46.
XX
KW   Human; enzyme; HIPHUM46; serine protease; gene therapy; osteoarthritis;
KW   serine protease activity modulation; dipeptidyl peptidase activity;
KW   musculoskeletal disease; Hepatitis B virus infection; myotonic dystrophy;
KW   Alzheimer's disease; parasupranuclear palsy; Huntington's disease;
KW   amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;
KW   irritable bowel syndrome; type I diabetes; faecal incontinence; anaemia;
KW   haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;
KW   colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;
KW   multiple sclerosis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          259..260
FT                   /note= "Paired glutamates of the beta propeller domain"
FT   Active-site     739
FT                   /label= Catalytic_serine_residue
FT   Active-site     817
FT                   /label= Catalytic_aspartate_residue
FT   Active-site     849
FT                   /label= Catalytic_histidine_residue
XX
PN   GB2374869-A.
XX
PD   30-OCT-2002.
XX
PF   22-JAN-2002; 2002GB-0001404.
XX
PR   23-JAN-2001; 2001GB-0001760.
XX
PA   (GLAX ) GLAXO GROUP LTD.
XX
PI   Edbrooke MR,  Lewis AP;
XX
DR   WPI; 2003-150703/15.
DR   N-PSDB; ABX12255.
XX
PT   Identifying modulators of serine protease activity useful for treating
PT   musculoskeletal diseases, by contacting cell expressing a novel serine
PT   protease polypeptide with a compound and monitoring serine protease
PT   activity  -
XX
PS   Claim 10; Page 26-29; 38pp; English.
XX
CC   The invention relates to a method of identifying a substance that
CC   modulates serine protease activity, comprising contacting a cell such as
CC   a neuronal cell, lung cell, intestinal cell or a cell infected with a
CC   virus, expressing a serine protease polypeptide (HIPHUM 46), or its
CC   variant having dipeptidyl peptidase activity, or a serine protease
CC   isolated from the cell with a test substance and monitoring for serine
CC   protease activity. The method is useful for identifying a substance that
CC   modulates serine protease activity. A modulator of the serine protease is
CC   useful in the manufacture of a medicament for treatment or prophylaxis of
CC   a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus
CC   infection, Alzheimer's disease, parasupranuclear palsy, myotonic
CC   dystrophy, Huntington's disease or amyotrophic lateral sclerosis.
CC   Additional disease that may be treated using modulators of the serine
CC   protease include malabsorption syndromes, irritable bowel syndrome, lung
CC   disease, type I diabetes, faecal incontinence, haemorrhoids, proctitis,
CC   rectal polyps, small bowel tumours, colorectal tumours, anaemia,
CC   dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple
CC   sclerosis. The present sequence represents the amino acid sequence of the
CC   human serine protease HIPHUM46.
XX
SQ   Sequence   882 AA;

  Query Match            100.0%;  Score 4700;  DB 24;  Length 882;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  882;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660

Qy        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720

Qy        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780

Qy        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840

Qy        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
                ||||||||||||||||||||||||||||||||||||||||||
Db        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882


RESULT 7
ABB97361
ID   ABB97361 standard; Protein; 782 AA.
XX
AC   ABB97361;
XX
DT   27-JUN-2002  (first entry)
XX
DE   Novel human protein SEQ ID NO: 629.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag.
XX
OS   Homo sapiens.
XX
PN   WO200222660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US26015.
XX
PR   11-SEP-2000; 2000US-0659671.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   N-PSDB; ABN32547.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis -
XX
PS   Example 2; SEQ ID NO 629; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC   stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC   e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a protein of the invention.
XX
SQ   Sequence   782 AA;

  Query Match             87.6%;  Score 4118;  DB 23;  Length 782;
  Best Local Similarity   88.7%;  Pred. No. 0;
  Matches  782;  Conservative    0;  Mismatches    0;  Indels  100;  Gaps    1;

Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQ--- 657

Qy        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        658 ------------------------------------------------------------ 657

Qy        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780
                                                    |||||||||||||||||||||||
Db        658 ------------------------------------VAIAGAPVTLWIFYDTGYTERYM 680

Qy        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 740

Qy        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
              |||||||||||||||||||||||||||||||||||||||||||
Db        741 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 782


RESULT 8
ABB97362
ID   ABB97362 standard; Protein; 724 AA.
XX
AC   ABB97362;
XX
DT   27-JUN-2002  (first entry)
XX
DE   Novel human protein SEQ ID NO: 630.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag.
XX
OS   Homo sapiens.
XX
PN   WO200222660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US26015.
```

```
XX
PR   11-SEP-2000; 2000US-0659671.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   N-PSDB; ABN32548.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis -
XX
PS   Example 2; SEQ ID NO 630; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC   stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC   e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a protein of the invention.
XX
SQ   Sequence   724 AA;

  Query Match             80.2%;  Score 3771;  DB 23;  Length 724;
  Best Local Similarity   82.1%;  Pred. No. 0;
  Matches  724;  Conservative    0;  Mismatches    0;  Indels  158;  Gaps    2;

Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||
Db        121 DLFQ-------------------------------------------------------- 124

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 --QQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 182

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 242

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 302

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 362

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 422

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 482

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 542

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQ--- 599

Qy        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720

Db        600 ------------------------------------------------------------ 599

Qy        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780
                                                   |||||||||||||||||||||||
Db        600 -------------------------------------VAIAGAPVTLWIFYDTGYTERYM 622

Qy        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        623 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 682

Qy        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
              ||||||||||||||||||||||||||||||||||||||||||
Db        683 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 724


RESULT 9
ABG61600
ID   ABG61600 standard; Protein; 658 AA.
XX
AC   ABG61600;
XX
DT   12-AUG-2002  (first entry)
XX
DE   Human DPRP-1 splice variant #7.
XX
KW   Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW   DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW   diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW   heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW   ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW   dyskinesia; reproductive disorder; inflammatory disorder;
KW   metabolic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200231134-A2.
XX
PD   18-APR-2002.
XX
PF   12-OCT-2001; 2001WO-US31874.
XX
PR   12-OCT-2000; 2000US-240117P.
XX
PA   (FERR ) FERRING BV.
XX
PI   Qi S,  Akinsanya KO,  Riviere PJ,  Junien J;
XX
DR   WPI; 2002-444178/47.
DR   N-PSDB; ABK83331.
XX
PT   New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT   the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT   viral infections, cancers, allergies, neurological disorders, or pain
PT   -
XX
PS   Disclosure; Page 70-72; 113pp; English.
XX
CC   The present invention relates to the isolation of novel human serine
CC   proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC   proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC   and nucleic acids encoding them are useful for treating infections
CC   such as fungal, bacterial, protozoan and viral infections, particularly
CC   infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC   pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC   bulimia, Parkinson's disease, acute heart failure, hypotension,
CC   hypertension, urinary retention, osteoporosis, angina pectoris,
CC   stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
```

```
CC   psychotic and neurological disorders (e.g. anxiety, dementia, or
CC   schizophrenia), and dyskinesias. These may also be used in discovering
CC   therapeutic agents for the treatment of reproductive, inflammatory and
CC   metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ   Sequence   658 AA;

  Query Match             74.6%;  Score 3504;  DB 23;  Length 658;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  655;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 655
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 655


RESULT 10
ABG61596
ID   ABG61596 standard; Protein; 661 AA.
XX
AC   ABG61596;
XX
DT   12-AUG-2002  (first entry)
XX
DE   Human DPRP-1 splice variant #3.
XX
KW   Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW   DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW   diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW   heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW   ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW   dyskinesia; reproductive disorder; inflammatory disorder;
KW   metabolic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200231134-A2.
XX
PD   18-APR-2002.
XX
PF   12-OCT-2001; 2001WO-US31874.
XX
PR   12-OCT-2000; 2000US-240117P.
XX
PA   (FERR ) FERRING BV.
XX
PI   Qi S,  Akinsanya KO,  Riviere PJ,  Junien J;
XX
DR   WPI; 2002-444178/47.
DR   N-PSDB; ABK83327.
XX
PT   New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT   the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT   viral infections, cancers, allergies, neurological disorders, or pain
PT   -
XX
PS   Disclosure; Page 63-65; 113pp; English.
XX
CC   The present invention relates to the isolation of novel human serine
CC   proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC   proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC   and nucleic acids encoding them are useful for treating infections
CC   such as fungal, bacterial, protozoan and viral infections, particularly
CC   infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC   pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC   bulimia, Parkinson's disease, acute heart failure, hypotension,
CC   hypertension, urinary retention, osteoporosis, angina pectoris,
CC   stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC   psychotic and neurological disorders (e.g. anxiety, dementia, or
CC   schizophrenia), and dyskinesias. These may also be used in discovering
CC   therapeutic agents for the treatment of reproductive, inflammatory and
CC   metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ   Sequence   661 AA;

  Query Match             74.6%;  Score 3504;  DB 23;  Length 661;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  655;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
```

```
Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 655
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 655
```

RESULT 11
ABG61594
ID ABG61594 standard; Protein; 690 AA.
XX
AC ABG61594;
XX
DT 12-AUG-2002 (first entry)
XX
DE Human DPRP-1 splice variant #1.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder.
XX
OS Homo sapiens.
XX
PN WO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI; 2002-444178/47.
DR N-PSDB; ABK83325.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT -
XX
PS Disclosure; Page 59-61; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ Sequence 690 AA;

```
  Query Match             74.6%;  Score 3504;  DB 23;  Length 690;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  655;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 655
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 655
```

RESULT 12
AAB93565
ID AAB93565 standard; Protein; 632 AA.
XX
AC AAB93565;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12964.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX

PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 12964; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 632 AA;

Query Match 70.9%; Score 3333.5; DB 22; Length 632;
Best Local Similarity 92.4%; Pred. No. 1.7e-312;
Matches 631; Conservative 0; Mismatches 1; Indels 51; Gaps 1;

```
Qy        200 MDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQE 259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQE 60

Qy        260 EFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTG 319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTG 120

Qy        320 TANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILL 379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILL 180

Qy        380 DRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHV 439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHV 240

Qy        440 FPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAIT 499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAIT 300

Qy        500 SGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSC 559
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        301 SGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRSYSHSC 360

Qy        560 CISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEI 619
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEI 420

Qy        620 FSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASL 679
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 FSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASL 480

Qy        680 GYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWS 739
              |||||||||||||||||||||||||||
Db        481 GYVVVVIDNRGSCHRGLKFEGAFKYKM------------------------------- 507

Qy        740 YGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQA 799
                                  ||||||||||||||||||||||||||||||||||||||
Db        508 ------------------VAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQA 549

Qy        800 EKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEH 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        550 EKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEH 609

Qy        860 YELHLLHYLQENLGSRIAALKVI 882
              |||||||||||||||||||||||
Db        610 YELHLLHYLQENLGSRIAALKVI 632
```

RESULT 13
ABG61601
ID ABG61601 standard; Protein; 613 AA.
XX
AC ABG61601;
XX
DT 12-AUG-2002 (first entry)
XX
DE Human DPRP-1 splice variant #8.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder.
XX
OS Homo sapiens.
XX
PN WO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI; 2002-444178/47.
DR N-PSDB; ABK83332.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT -
XX
PS Disclosure; Page 73-75; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),

```
CC   pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC   bulimia, Parkinson's disease, acute heart failure, hypotension,
CC   hypertension, urinary retention, osteoporosis, angina pectoris,
CC   stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC   psychotic and neurological disorders (e.g. anxiety, dementia, or
CC   schizophrenia), and dyskinesias. These may also be used in discovering
CC   therapeutic agents for the treatment of reproductive, inflammatory and
CC   metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ   Sequence   613 AA;

  Query Match             68.9%;  Score 3236;  DB 23;  Length 613;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-303;
  Matches  607;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWININHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWININHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDS 607
              |||||||
Db        601 WATILDS 607


RESULT 14
ABG61592
ID   ABG61592 standard; Protein; 863 AA.
XX
AC   ABG61592;
XX
DT   12-AUG-2002  (first entry)
XX
DE   Human DPPIV related serine protease DPRP-2.
XX
KW   Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW   DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW   diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW   heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW   ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW   dyskinesia; reproductive disorder; inflammatory disorder;
KW   metabolic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200231134-A2.
XX
PD   18-APR-2002.
XX
PF   12-OCT-2001; 2001WO-US31874.
XX
PR   12-OCT-2000; 2000US-240117P.
XX
PA   (FERR ) FERRING BV.
XX
PI   Qi S,  Akinsanya KO,  Riviere PJ,  Junien J;
XX
DR   WPI; 2002-444178/47.
DR   N-PSDB; ABK83323.
XX
PT   New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT   the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT   viral infections, cancers, allergies, neurological disorders, or pain
PT   -
XX
PS   Claim 17; Fig 1; 113pp; English.
XX
CC   The present invention relates to the isolation of novel human serine
CC   proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC   proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC   and nucleic acids encoding them are useful for treating infections
CC   such as fungal, bacterial, protozoan and viral infections, particularly
CC   infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC   pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC   bulimia, Parkinson's disease, acute heart failure, hypotension,
CC   hypertension, urinary retention, osteoporosis, angina pectoris,
CC   stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC   psychotic and neurological disorders (e.g. anxiety, dementia, or
CC   schizophrenia), and dyskinesias. These may also be used in discovering
CC   therapeutic agents for the treatment of reproductive, inflammatory and
CC   metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ   Sequence   863 AA;

  Query Match             61.1%;  Score 2870;  DB 23;  Length 863;
  Best Local Similarity   61.5%;  Pred. No. 1.8e-267;
  Matches  517;  Conservative  134;  Mismatches  187;  Indels    2;  Gaps    2;

Qy         35 FYVERYSWSQLKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRE 94
              | |:::||  |: ::  :||| | :: |||||| ||:: |  |||| |:||| |   :||
Db         24 FQVQKHSWDGLRSIIHGSRKYSGLIVNKAPHDFQFVQKTDESGPHSHRLYYLGMPYGSRE 83

Qy         95 NTLFYSEIPKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYD 154
              |:| |||||| : : |:|:|||| :|| ||||  :|:|||||||||||||:|  || |||
Db         84 NSLLYSEIPKKVRKEALLLLSWKQMLDHFQATPHHGVYSREEELLRERKRLGVFGITSYD 143

Qy        155 YHQGSGTFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAF 214
              :|  || ||||| : ::| :|||  ||   |::|  ::| |   |||||:||||| : :|
Db        144 FHSESGLFLFQASNSLFHCRDGGKNGFMVSPMKPLEIKTQCSGPRMDPKICPADPAFFSF 203

Qy        215 IHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAE 274
              |:::|:|::|| | ||||||: |  |:|: :| :||||||||:||||||::|||||| |
Db        204 INNSDLWVANIETGEERRLTFCHQGLSNVLDDPKSAGVATFVIQEEFDRFTGYWWCPTAS 263

Qy        275 TTPSGG-KILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIM 333
                | | | ||||||| ||||||:||| || || |: ||:|||:||: |||:  |::|
Db        264 WEGSEGLKTLRILYEEVDESEVEVIHVPSPALEERKTDSYRYPRTGSKNPKIALKLAEFQ 323

Qy        334 IDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLISP 393
              |::|:|:   :|||:|||  ||  |||||||||| :|||||:: ||| |  ||:||: |
Db        324 TDSQGKIVSTQEKELVQPFSSLFPKVEYIARAGWTRDGKYAWAMFLDRPQQWLQLVLLPP 383
```

```
Qy        394 ELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWININHDIFHVFPQSH-EEEIEFIF 452
Db        384 ALFIPSTENEEQRLASARAVPRNVQPYVVYEEVTNVWINVHDIFYPFPQSEGEDELCFLR 443

Qy        453 ASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSN 512
Db        444 ANECKTGFCHLYKVTAVLKSQGYDWSEPFSPGEDEFKCPIKEEIALTSGEWEVLARHGSK 503

Qy        513 IQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKY 572
Db        504 IWVNEETKLVYFQGTKDTPLEHHLYVVSYEAAGEIVRLTTPGFSHSCSMSQNFDMFVSHY 563

Qy        573 SNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYG 632
Db        564 SSVSTPPCVHVYKLSGPDDDPLHKQPRFWASMMEAASCPPDYVPPEIFHFHTRSDVRLYG 623

Qy        633 MLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSC 692
Db        624 MIYKPHALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNTLASLGYAVVVIDGRGSC 683

Qy        693 HRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQR 752
Db        684 QRGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAIHGWSYGGFLSLMGLIHK 743

Qy        753 SDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLL 812
Db        744 PQVFKVAIAGAPVTVWMAYDTGYTERYMDVPENNQHGYEAGSVALHVEKLPNEPNRLLIL 803

Qy        813 HGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQENL 872
Db        804 HGFLDENVHFFHTNFLVSQLIRAGKPYQLQIYPNERHSIRCPESGEHYEVTLLHFLQEYL 863
```

RESULT 15
ABB98134
ID ABB98134 standard; Protein; 892 AA.
XX
AC ABB98134;
XX
DT 17-OCT-2002 (first entry)
XX
DE Human PMMM Incyte ID 7972712CD1.
XX
KW Human; PMMM; protein modification and maintenance molecule;
KW anticonvulsant; neuroprotective; nootropic; cytostatic; antipsoriatic;
KW antiasthmatic; dermatological; antidiabetic; antiparkinsonian;
KW antianaemic; antiinflammatory; antiulcer; antianginal; cardiant;
KW hepatotropic; osteopathic; antiemetic; antipyretic; virucide;
KW antibacterial; fungicide; gastrointestinal; antidiarrheic; laxative;
KW haemostatic; anti-HIV; thrombolytic; anticoagulant; gynaecological;
KW cardiovascular; antiarteriosclerotic; hypotensive; vasotropic;
KW antitumour; antirheumatic; immunosuppressive; antiallergic; antithyroid;
KW nephrotropic; antigout; thyromimetic; antiarthritic; uropathic;
KW ophthalmological; antiparasitic; tranquiliser; vulnerary; keratolytic;
KW auditory; antiseborrheic; antidepressant; neuroleptic; antiinfertility;
KW anthelmintic; protozoacide; Crohn's disease; hypertension;
KW autoimmune; inflammatory; anaemia; cell proliferative; developmental;
KW epithelial; scabies; neurological; Alzheimer's disease; reproductive;
KW ectopic pregnancy; gene therapy; vaccine; disorder;
KW dipeptidyl peptidase 8.
XX
OS Homo sapiens.
XX
PN WO200246383-A2.
XX
PD 13-JUN-2002.
XX
PF 05-DEC-2001; 2001WO-US46964.
XX
PR 08-DEC-2000; 2000US-254399P.
PR 21-DEC-2000; 2000US-257803P.
PR 05-JAN-2001; 2001US-260110P.
PR 19-JAN-2001; 2001US-262851P.
PR 25-JAN-2001; 2001US-264623P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Azimzai Y, Kallick DA, Baughn MR, Griffin JA, Swarnakar A;
PI Lal PG, Walia NK, Hafalia AJA, Gandhi AR, Au-Young J, Elliott VS;
PI Ramkumar J, Thangavelu K, Lu Y, Warren BA, Lu DAM, Lee EA;
PI Tribouley CM, Arvizu C, Delegeane AM, Yao MG, Khan FA;
PI Sanjanwala MM;
XX
DR WPI; 2002-519664/55.
DR N-PSDB; ABQ75955.
XX
PT New isolated Protein Modification and Maintenance polypeptides, useful
PT for diagnosis, and treatment of e.g. gastrointestinal disorders -
XX
PS Claim 1 (a); Page 172-174; 200pp; English.
XX
CC The invention relates to an isolated Protein Modification and Maintenance
CC (PMM) polypeptide. Polypeptides of the invention may be used in the
CC diagnosis, treatment and prevention of disorders associated with
CC decreased expression or activity of PMMM. These include gastrointestinal
CC disorders (e.g. Crohn's disease), cardiovascular disorders (e.g.
CC hypertension), autoimmune/inflammatory disorders (e.g. anaemia), cell
CC proliferative disorders, developmental disorders, epithelial disorders
CC (e.g. scabies), neurological disorders (e.g. Alzheimer's disease)
CC reproductive disorders (e.g. ectopic pregnancy), and in gene therapy or a
CC vaccine for such diseases. They may also be used in the assessment of the
CC effects of exogenous compound on the expression of nucleic acid and amino
CC acid sequences of protein modification and maintenance molecules. The
CC current sequence represents a human PMMM of the invention, which has been
CC found to have homology with human dipeptidyl peptidase 8.
XX
SQ Sequence 892 AA;

Query Match 61.1%; Score 2870; DB 23; Length 892;
Best Local Similarity 61.5%; Pred. No. 1.9e-267;
Matches 517; Conservative 134; Mismatches 187; Indels 2; Gaps 2;

```
Qy         35 FYVERYSWSQLKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRE 94
Db         53 FQVQKHSWDGLRSIIHGSRKYSGLIVNKAPHDFQFVQKTDESGPHSHRLYYLGMPYGSRE 112

Qy         95 NTLFYSEIPKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYD 154
Db        113 NSLLYSEIPKKVRKEALLLLSWKQMLDHFQATPHHGVYSREEELLRERKRLGVFGITSYD 172

Qy        155 YHQGSGTFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAF 214
Db        173 FHSESGLFLFQASNSLFHCRDGGKNGFMVSPMKPLEIKTQCSGPRMDPKICPADPAFFSF 232

Qy        215 IHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAE 274
Db        233 INNSDLWVANIETGEERRLTFCHQGLSNVLDDPKSAGVATFVIQEEFDRFTGYWWCPTAS 292

Qy        275 TTPSGG-KILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIM 333
Db        293 WEGSEGLKTLRILYEEVDESEVEVIHVPSPALEERKTDSYRYPRTGSKNPKIALKLAEFQ 352

Qy        334 IDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLISP 393
Db        353 TDSQGKIVSTQEKELVQPFSSLFPKVEYIARAGWTRDGKYAWAMFLDRPQQWLQLVLLPP 412

Qy        394 ELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWININHDIFHVFPQSH-EEEIEFIF 452
Db        413 ALFIPSTENEEQRLASARAVPRNVQPYVVYEEVTNVWINVHDIFYPFPQSEGEDELCFLR 472

Qy        453 ASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSN 512
Db        473 ANECKTGFCHLYKVTAVLKSQGYDWSEPFSPGEDEFKCPIKEEIALTSGEWEVLARHGSK 532

Qy        513 IQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKY 572
```

```
                 | |:|  :||||:||||:||||||||||||   ||: |||  |:|||| :||: | |:| |
Db        533 IWVNEETKLVYFQGTKDTPLEHHLYVVSYEAAGEIVRLTTPGFSHSCSMSQNFDMFVSHY 592

Qy        573 SNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYG 632
                 |:   | || :|||| |:|||  |   |||:::::|   ||| ||||| | : :   |||
Db        593 SSVSTPPCVHVYKLSGPDDDPLHKQPRFWASMMEAASCPPDYVPPEIFHFHTRSDVRLYG 652

Qy        633 MLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSC 692
                 |:|||| ||||||:||||||:|||||||||||| |||:|| |||||||||| ||||| ||||
Db        653 MIYKPHALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNTLASLGYAVVVIDGRGSC 712

Qy        693 HRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQR 752
                  |||:|||| | :|||:||:|||||||::| :| |||| || |||||||:|||| |: :
Db        713 QRGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAIHGWSYGGFLSLMGLIHK 772

Qy        753 SDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLL 812
                :|:|||||||||:|: |||||||||  |: |: ||  ||||:  || |:|||||:|
Db        773 PQVFKVAIAGAPVTVWMAYDTGYTERYMDVPENNQHGYEAGSVALHVEKLPNEPNRLLIL 832

Qy        813 HGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQENL 872
                |||||||||| ||: |:| |:|||||| ||||| |||||| |||||||: |||:||| |
Db        833 HGFLDENVHFFHTNFLVSQLIRAGKPYQLQIYPNERHSIRCPESGEHYEVTLLHFLQEYL 892

Search completed: October 15, 2003, 17:10:43
Job time : 58.7858 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:10:56 ; Search time 19.6777 Seconds
(without alignments)
1896.467 Million cell updates/sec

Title: US-10-070-464-1
Perfect score: 4700
Sequence: 1 MAAAMETEQLGVEIFETADC.....HLLHYLQENLGSRIAALKVI 882

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1645.5	35.0	310	4	US-09-794-236-4	Sequence 4, Appli
2	582.5	12.4	771	4	US-09-462-284-2	Sequence 2, Appli
3	529	11.3	766	4	US-10-002-593-6	Sequence 6, Appli
4	525	11.2	755	5	PCT-US93-07923-3	Sequence 3, Appli
5	525	11.2	759	5	PCT-US93-07923-2	Sequence 2, Appli
6	525	11.2	766	1	US-08-230-491A-3	Sequence 3, Appli
7	525	11.2	766	1	US-08-619-280A-3	Sequence 3, Appli
8	525	11.2	766	2	US-08-940-391-3	Sequence 3, Appli
9	525	11.2	766	4	US-09-794-236-1	Sequence 1, Appli
10	454	9.7	760	1	US-08-230-491A-2	Sequence 2, Appli
11	454	9.7	760	1	US-08-619-280A-2	Sequence 2, Appli
12	454	9.7	760	2	US-08-940-391-2	Sequence 2, Appli
13	282	6.0	632	3	US-09-016-080-1	Sequence 1, Appli
14	249.5	5.3	657	4	US-09-355-166-1	Sequence 1, Appli
15	225.5	4.8	593	5	PCT-US93-07923-11	Sequence 11, Appl
16	192	4.1	622	2	US-08-664-646A-2	Sequence 2, Appli
17	192	4.1	622	2	US-09-066-285-2	Sequence 2, Appli
18	192	4.1	622	3	US-09-261-006-2	Sequence 2, Appli
19	192	4.1	622	3	US-08-951-088-2	Sequence 2, Appli
20	192	4.1	622	4	US-09-609-566-2	Sequence 2, Appli
21	192	4.1	622	4	US-09-609-570-2	Sequence 2, Appli
22	192	4.1	622	4	US-09-427-372-2	Sequence 2, Appli
23	192	4.1	622	4	US-09-693-554-2	Sequence 2, Appli
24	180.5	3.8	721	4	US-09-390-234-20	Sequence 20, Appl
25	175.5	3.7	614	4	US-09-252-991A-20060	Sequence 20060, A
26	133	2.8	344	4	US-09-724-623-79	Sequence 79, Appl
27	132.5	2.8	275	4	US-09-198-452A-181	Sequence 181, App
28	123	2.6	1022	1	US-08-271-364A-8	Sequence 8, Appli
29	123	2.6	1022	2	US-08-222-715B-27	Sequence 27, Appl
30	120.5	2.6	686	3	US-09-368-169-8	Sequence 8, Appli
31	116	2.5	561	3	US-09-233-989-5	Sequence 5, Appli
32	113.5	2.4	1209	4	US-09-749-588-4	Sequence 4, Appli
33	112.5	2.4	529	4	US-09-252-991A-24711	Sequence 24711, A
34	112.5	2.4	1285	1	US-07-582-945-2	Sequence 2, Appli
35	112.5	2.4	1285	2	US-08-453-141-2	Sequence 2, Appli
36	112.5	2.4	1285	3	US-08-293-314-2	Sequence 2, Appli
37	110.5	2.4	836	4	US-09-491-356C-21	Sequence 21, Appl
38	110	2.3	1296	1	US-08-480-604A-28	Sequence 28, Appl
39	110	2.3	1296	2	US-08-405-496A-28	Sequence 28, Appl
40	110	2.3	1296	3	US-08-915-136-28	Sequence 28, Appl
41	109	2.3	570	3	US-09-068-960-13	Sequence 13, Appl
42	109	2.3	808	4	US-09-198-452A-327	Sequence 327, App
43	108.5	2.3	782	4	US-09-585-858-29	Sequence 29, Appl
44	106	2.3	256	4	US-09-355-166-2	Sequence 2, Appli
45	106	2.3	1362	4	US-09-134-001C-4955	Sequence 4955, Ap

ALIGNMENTS

RESULT 1
US-09-794-236-4
; Sequence 4, Application US/09794236
; Patent No. 6337069
; GENERAL INFORMATION:
; APPLICANT: Grouzmann, Eric
; APPLICANT: Lacroix, Jean-Silvain
; APPLICANT: Monod, Michel
; TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
; FILE REFERENCE: 81985/276823
; CURRENT APPLICATION NUMBER: US/09/794,236
; CURRENT FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-794-236-4

Query Match 35.0%; Score 1645.5; DB 4; Length 310;
Best Local Similarity 86.4%; Pred. No. 5.2e-159;
Matches 310; Conservative 0; Mismatches 0; Indels 49; Gaps 1;

```
Qy        524 FEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSL 583
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 FEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSL 60

Qy        584 YKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSPESTTGFTLYGMLYKPHDLQPG 643
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSPESTTGFTLYGMLYKPHDLQPG 120

Qy        644 KKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFK 703
              ||||||||||||||
Db        121 KKYPTVLFIYGGPQ---------------------------------------------- 134

Qy        704 YKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGA 763
                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 ---GQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGA 191

Qy        764 PVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFA 823
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 PVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFA 251

Qy        824 HTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 HTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 310
```

RESULT 2

```
US-09-462-284-2
; Sequence 2, Application US/09462284
; Patent No. 6309868
; GENERAL INFORMATION:
;  APPLICANT:  Nestec S.A.
;  APPLICANT:  Monod, Michel
;  APPLICANT:  Doumas, Agnes
;  APPLICANT:  Affolter, Micheal
;  APPLICANT:  Van Den Broek, Peter
;  TITLE OF INVENTION:  CLONING OF THE
;  TITLE OF INVENTION:  PROLYL-DIPEPTIDYL-PEPTIDASE FROM
;  TITLE OF INVENTION:  ASPERGILLUS ORYZAE
;  FILE REFERENCE:  8265-298
;  CURRENT APPLICATION NUMBER: US/09/462,284
;  CURRENT FILING DATE:  2000-01-03
;  NUMBER OF SEQ ID NOS:  9
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 771
;   TYPE: PRT
;   ORGANISM: Fungus
US-09-462-284-2

  Query Match             12.4%;  Score 582.5;  DB 4;  Length 771;
  Best Local Similarity   28.4%;  Pred. No. 7.6e-50;
  Matches  198;  Conservative  103;  Mismatches  262;  Indels  135;  Gaps   30;

Qy        203 KLCPADPD---------W------IAFIHSND--IWISNIVTREERRLTYVHNELANMEE 245
Db        136 KSVPLMPDQEGDIQYAQWSPVGNTIAFVRENDLYVWDNGTVTR--------------ITD 181

Qy        246 DA---RSAGVATFVLQEEF--DRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
Db        182 DGGPDMFHGVPDWIYEEEILGDRYA-LWFSPDGE---------YLAYLSFNETGVPTYTV 231

Qy        301 TSPMLETRRADSF------RYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEI 354
Db        232 QYYMDNQEIAPAYPWELKIRYPKVSQTNPTVTLS-----------LLNIASKEVKQAPID 280

Qy        355 LFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVP 414
Db        281 AFESTDLII-------GEVAW--LTDTHTT------VAAKAFNRVQD--QQKVVAVDTAS 323

Qy        415 DSVTPLIIYEETTDIWI-NIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKES 473
Db        324 NKAT-VISDRDGTDGWLDNLLSMKYIGPIKPSDKDAYYIDISDHSGWAHLYLFP------ 376

Qy        474 KYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLE 533
Db        377 ----VSGGEPIP-----------LTKGDWEVT----SILSIDQERQLVYYLSTQHHSTE 416

Qy        534 HHLYVVSYVNPGEVTRLTDRGYS--HSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPED 591
Db        417 RHLYSVSY-STFAVTPLVDDTVAAYWSASFSANSGYYILTYGGPDVPY-QELYTTNS--- 471

Qy        592 DPTCKTKEFWATILDSAGPL---PDYTPPEIFSFE--STTGFTLYGMLYKPHDLQPGKKY 646
Db        472 -----TKPL-RTITDNAKVLEQIKDYALPNITYFELPLPSGETLNVMQRLPPGFSPDKKY 525

Qy        647 PTVLFIYGGPQVQLVNNRFKGVKY-FRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYK 705
Db        526 PILFTPYGGPGAQEVTKRWQALNFKAYVASDSELEYVTWTVDNRGTGFKGRKFRSAVTRQ 585

Qy        706 MGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPV 765
Db        586 LGLLEAEDQIYAAQQ-AANIPWIDADHIGIWGWSFGGYLTSKVLEKDSGAFTLGVITAPV 644

Qy        766 TLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHT 825
Db        645 SDWRFYDSMYTERYMKTLSTNEEGYETSAV-RKTDGFKNVEGGFLIQHGTGDDNVHFQNS 703

Qy        826 SILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELH 863
Db        704 AALVDLLMGDG------VSPEKLHSQWFTDSDHGISYH 735

RESULT 3
US-10-002-593-6
; Sequence 6, Application US/10002593
; Patent No. 6586198
; GENERAL INFORMATION:
;  APPLICANT: Vanderbilt University
;  APPLICANT:  Brown, Nancy J.
;  TITLE OF INVENTION: BIOLOGICAL MARKERS AND DIAGNOSTIC TESTS FOR ANGIOTENSIN CONVERTI
;  TITLE OF INVENTION:  INHIBITOR AND VASOPEPTIDASE INHIBITOR ASSOCIATED ANGIOEDEMA
;  FILE REFERENCE: Atty Docket No. 6586198 1242/48/2
;  CURRENT APPLICATION NUMBER: US/10/002,593
;  CURRENT FILING DATE:  2001-10-31
;  PRIOR APPLICATION NUMBER: 60/244,524
;  PRIOR FILING DATE: 2000-10-31
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 766
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-002-593-6

  Query Match             11.3%;  Score 529;  DB 4;  Length 766;
  Best Local Similarity   26.5%;  Pred. No. 2.1e-44;
  Matches  186;  Conservative  109;  Mismatches  280;  Indels  128;  Gaps   32;

Qy        206 PADPDWI---------AFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFV 256
Db        149 PNNTQWVTWSPVGHKLAYVWNNDIYVKIEPNLPSYRITWTG------KEDIIYNGITDWV 202

Qy        257 LQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH---VTSPMLETRRADS 312
Db        203 YEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSFYSDESLQYPKTVR 253

Qy        313 FRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGK 372
Db        254 VPYPKAGAVNPTVKFFV--VNTDSLSSVTNATSIQITAPASMLI-GDHYLCDVTWA---- 306

Qy        373 YAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWIN 432
Db        307 ---------TQERISLQWL---------------RRIQNY--SVMDICDYDESSGRW-N 338

Qy        433 IHDIFHVFPQSHEEEIEFIFASECKTGFRH-------LYKITSILKESKYKRSSGGLPAP 485
Db        339 C-----LVARQHIEMSTTGWVGRFRPSEPHFTLDGNSFYKIIS--NEEGYRHI------- 384

Qy        486 SDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGT-KDSPLEHHLYVVSYVNP 544
Db        385 CYFQIDKKDCTFITKGTWEVIG-----IEALTSDYLYYISNEYKGMPGGRNLYKIQLIDY 439

Qy        545 GEVTRLTDRGYSHSCCIS-QHCDFFISKYSNQKNPH---C----VSLYKLSSPEDDPTCK 596
Db        440 TKVTCL-------SCELNPERCQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVNDKGLR 492

Qy        597 TKEFWATILDSAGPLPDYTPPE---IFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIY 653
Db        493 VLED-NSALDKM--LQNVQMPSKKLDFIILNETKF-WYQMILPPH-FDKSKKYPLLLDVY 547

Qy        654 GGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLKFEGAFKYKMGQI 709
Db        548 AGPCSQ------KADTVFRLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRRLGTF 601

Qy        710 EIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWI 769
Db        602 EVEDQIEAARQF-SKMGFVDNKRIAIWGWSYGGYVTSMVLGSGSGVFKCGIAVAPVSRWE 660

Qy        770 FYDTGYTERYMG--HPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSI 827
Db        661 YYDSVYTERYMGLPTPEDNLDHYRNSTVMSRAENF--KQVEYLLIHGTADDNVHFQQSAQ 718
```

```
Qy        828 LLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQE 870
Db        719 ISKALVDVGVDFQAMWYTDEDHGIASSTAHQHIYTHMSHFIKQ 761
```

RESULT 4
PCT-US93-07923-3
; Sequence 3, Application PC/TUS9307923
; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao
; APPLICANT: Schlossman, Stuart F.
; APPLICANT: Tanaka, Toshiaki
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07923
; FILING DATE: 19930819
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/934,162
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: 07/832,211
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/055002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 755
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
PCT-US93-07923-3

Query Match 11.2%; Score 525; DB 5; Length 755;
Best Local Similarity 26.3%; Pred. No. 5.4e-44;
Matches 186; Conservative 108; Mismatches 277; Indels 136; Gaps 33;

```
Qy        206 PADPDWI---------AFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFV 256
Db        138 PNNTQWVTWSPVGHKLAYVWNNDIYVKIEPNLPSYRITWTG------KEDIIYNGITDWV 191

Qy        257 LQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH---VTSPMLETRRADS 312
Db        192 YEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSFYSDESLQYPKTVR 242

Qy        313 FRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGK 372
Db        243 VPYPKAGAVNPTVKFFV--VNTDSLSSVTNATSIQITAPASMLI-GDHYLCDVTWA---- 295

Qy        373 YAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWIN 432
Db        296 ---------TQERISLQWL---------------RRIQNY--SVMDICDYDESSGRW-N 327

Qy        433 IHDIFHVFPQSHEEEIEFIFASECKTGFRH-------LYKITSILKESKYKRSSGGLPAP 485
```

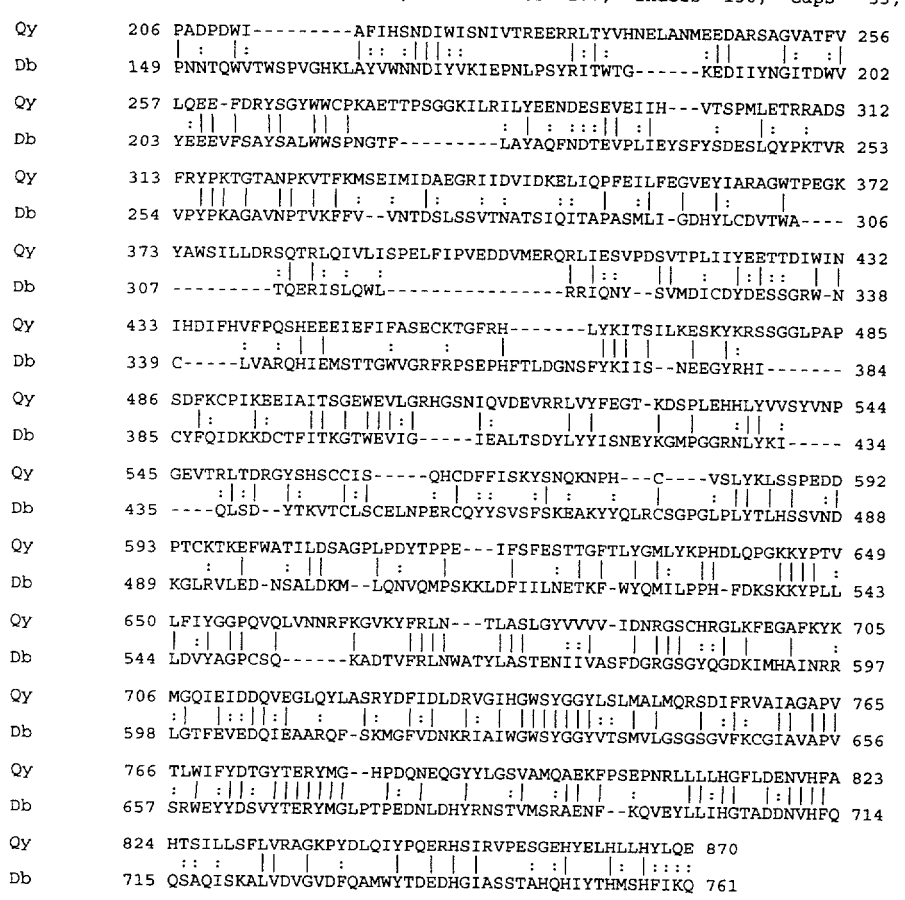
```
Db        328 C-----LVARQHIEMSTTGWVGRFRPSEPHFTLDGNSFYKIIS--NEEGYRHI------- 373

Qy        486 SDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGT-KDSPLEHHLYVVSYVNP 544
Db        374 CYFQIDKKDCTFITKGTWEVIG-----IEALTSDYLYYISNEYKGMPGGRNLYKI----- 423

Qy        545 GEVTRLTDRGYSHSCCIS-----QHCDFFISKYSNQKNPH---C----VSLYKLSSPEDD 592
Db        424 ----QLSD--YTKVTCLSCELNPERCQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVND 477

Qy        593 PTCKTKEFWATILDSAGPLPDYTPPE---IFSFESTTGFTLYGMLYKPHDLQPGKKYPTV 649
Db        478 KGLRVLED-NSALDKM--LQNVQMPSKKLDFIILNETKF-WYQMILPPH-FDKSKKYPLL 532

Qy        650 LFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLKFEGAFKYK 705
Db        533 LDVYAGPCSQ------KADTVFRLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRR 586

Qy        706 MGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPV 765
Db        587 LGTFEVEDQIEAARQF-SKMGFVDNKRIAIWGWSYGGYVTSMVLGSGSGVFKCGIAVAPV 645

Qy        766 TLWIFYDTGYTERYMG--HPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFA 823
Db        646 SRWEYYDSVYTERYMGLPTPEDNLDHYRNSTVMSRAENF--KQVEYLLIHGTADDNVHFQ 703

Qy        824 HTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQE 870
Db        704 QSAQISKALVDVGVDFQAMWYTDEDHGIASSTAHQHIYTHMSHFIKQ 750
```

RESULT 5
PCT-US93-07923-2
; Sequence 2, Application PC/TUS9307923
; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao
; APPLICANT: Schlossman, Stuart F.
; APPLICANT: Tanaka, Toshiaki
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07923
; FILING DATE: 19930819
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/934,162
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: 07/832,211
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/055002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 759
; TYPE: amino acid

```
;      STRANDEDNESS:
;      TOPOLOGY:  linear
PCT-US93-07923-2

  Query Match             11.2%;  Score 525;  DB 5;  Length 759;
  Best Local Similarity   26.3%;  Pred. No. 5.4e-44;
  Matches  186;  Conservative  108;  Mismatches  277;  Indels  136;  Gaps   33;

Qy        206 PADPDWI---------AFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFV 256
              |  :  |:         |:: :|||::         |:|:        :||    |:  :|
Db        142 PNNTQWVTWSPVGHKLAYVWNNDIYVKIEPNLPSYRITWTG------KEDIIYNGITDWV 195

Qy        257 LQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH---VTSPMLETRRADS 312
               :|| |  ||  || |              : | : :::|| :|     :   |:  :
Db        196 YEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSFYSDESLQYPKTVR 246

Qy        313 FRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGK 372
              ||| |  || | | :  :  |:   : :    ::  |  :|  |  |:    |
Db        247 VPYPKAGAVNPTVKFFV--VNTDSLSSVTNATSIQITAPASMLI-GDHYLCDVTWA---- 299

Qy        373 YAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWIN 432
                  :| |: :  :                  | |::   ||  :  |:|::  | |
Db        300 ---------TQERISLQWL---------------RRIQNY--SVMDICDYDESSGRW-N 331

Qy        433 IHDIFHVFPQSHEEEIEFIFASECKTGFRH-------LYKITSILKESKYKRSSGGLPAP 485
                  :  | |      :    |          ||| |    |  |:
Db        332 C-----LVARQHIEMSTTGWVGRFRPSEPHFTLDGNSFYKIIS--NEEGYRHI------- 377

Qy        486 SDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGT-KDSPLEHHLYVVSYVNP 544
              |:   |:   || | |||:|     |:      | |      |  |   :|| :
Db        378 CYFQIDKKDCTFITKGTWEVIG-----IEALTSDYLYYISNEYKGMPGGRNLYKI----- 427

Qy        545 GEVTRLTDRGYSHSCCIS-----QHCDFFISKYSNQKNPH---C----VSLYKLSSPEDD 592
                :|:|  |:   |:|     : | ::   :| :   :   |    : || | |  :|
Db        428 ----QLSD--YTKVTCLSCELNPERCQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVND 481

Qy        593 PTCKTKEFWATILDSAGPLPDYTPPE---IFSFESTTGFTLYGMLYKPHDLQPGKKYPTV 649
                :  |   : ||    | :   |      |   : | |  | |:  ||      |||| :
Db        482 KGLRVLED-NSALDKM--LQNVQMPSKKLDFIILNETKF-WYQMILPPH-FDKSKKYPLL 536

Qy        650 LFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLKFEGAFKYK 705
              | :| ||  |      |    ||||    |||   ::|   | ||| ::| |   |   :
Db        537 LDVYAGPCSQ------KADTVFRLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRR 590

Qy        706 MGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPV 765
              :|  |::||:|  :   |:  |:|  |: | |||||||:: | |   | :|:  || |||
Db        591 LGTFEVEDQIEAARQF-SKMGFVDNKRIAIWGWSYGGYVTSMVLGSGSGVFKCGIAVAPV 649

Qy        766 TLWIFYDTGYTERYMG--HPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFA 823
              : | :||: |||||||   |: |   |   :|  :|| |  :    ||:||  |:||||
Db        650 SRWEYYDSVYTERYMGLPTPEDNLDHYRNSTVMSRAENF--KQVEYLLIHGTADDNVHFQ 707

Qy        824 HTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLLHYLQE 870
                :: :   ||  |  :    |  | | |    : :|   |: |::::
Db        708 QSAQISKALVDVGVDFQAMWYTDEDHGIASSTAHQHIYTHMSHFIKQ 754


RESULT 6
US-08-230-491A-3
; Sequence 3, Application US/08230491A
; Patent No. 5587299
;  GENERAL INFORMATION:
;    APPLICANT:  Rettig, Wolfgang J.; Scanlan, Matthew J.;
;    APPLICANT:  Garin-Chesa, Pilar; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULE CODING FOR
;    TITLE OF INVENTION:  FIBROBLAST ACTIVATION PROTEIN ` AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FELFE & LYNCH
;      STREET:  805 THIRD AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE 3.5 inch 1.2 MB STORAGE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  WORDPERFECT - ASC II
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/230,491A
;      FILING DATE:  20-APRIL-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5587299man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 330
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  766 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-230-491A-3

  Query Match             11.2%;  Score 525;  DB 1;  Length 766;
  Best Local Similarity   26.3%;  Pred. No. 5.5e-44;
  Matches  186;  Conservative  108;  Mismatches  277;  Indels  136;  Gaps   33;

Qy        206 PADPDWI---------AFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFV 256
              |  :  |:         |:: :|||::         |:|:        :||    |:  :|
Db        149 PNNTQWVTWSPVGHKLAYVWNNDIYVKIEPNLPSYRITWTG------KEDIIYNGITDWV 202

Qy        257 LQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH---VTSPMLETRRADS 312
               :|| |  ||  || |              : | : :::|| :|     :   |:  :
Db        203 YEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSFYSDESLQYPKTVR 253

Qy        313 FRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGK 372
              ||| |  || | | :  :  |:   : :    ::  |  :|  |  |:    |
Db        254 VPYPKAGAVNPTVKFFV--VNTDSLSSVTNATSIQITAPASMLI-GDHYLCDVTWA---- 306

Qy        373 YAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWIN 432
                  :| |: :  :                  | |::   ||  :  |:|::  | |
Db        307 ---------TQERISLQWL---------------RRIQNY--SVMDICDYDESSGRW-N 338

Qy        433 IHDIFHVFPQSHEEEIEFIFASECKTGFRH-------LYKITSILKESKYKRSSGGLPAP 485
                  :  : | |      :    |          ||| |    |  |:
Db        339 C-----LVARQHIEMSTTGWVGRFRPSEPHFTLDGNSFYKIIS--NEEGYRHI------- 384

Qy        486 SDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGT-KDSPLEHHLYVVSYVNP 544
              |:   |:   || | |||:|     |:      | |      |  |   :|| :
Db        385 CYFQIDKKDCTFITKGTWEVIG-----IEALTSDYLYYISNEYKGMPGGRNLYKI----- 434

Qy        545 GEVTRLTDRGYSHSCCIS-----QHCDFFISKYSNQKNPH---C----VSLYKLSSPEDD 592
                :|:|  |:   |:|     : | ::   :| :   :   |    : || | |  :|
Db        435 ----QLSD--YTKVTCLSCELNPERCQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVND 488

Qy        593 PTCKTKEFWATILDSAGPLPDYTPPE---IFSFESTTGFTLYGMLYKPHDLQPGKKYPTV 649
                :  |   : ||    | :   |      |   : | |  | |:  ||      |||| :
Db        489 KGLRVLED-NSALDKM--LQNVQMPSKKLDFIILNETKF-WYQMILPPH-FDKSKKYPLL 543

Qy        650 LFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLKFEGAFKYK 705
              | :| ||  |      |    ||||    |||   ::|   | ||| ::| |   |   :
Db        544 LDVYAGPCSQ------KADTVFRLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRR 597

Qy        706 MGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPV 765
              :|  |::||:|  :   |:  |:|  |: | |||||||:: | |   | :|:  || |||
Db        598 LGTFEVEDQIEAARQF-SKMGFVDNKRIAIWGWSYGGYVTSMVLGSGSGVFKCGIAVAPV 656

Qy        766 TLWIFYDTGYTERYMG--HPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFA 823
              : | :||: |||||||   |: |   |   :|  :|| |  :    ||:||  |:||||
Db        657 SRWEYYDSVYTERYMGLPTPEDNLDHYRNSTVMSRAENF--KQVEYLLIHGTADDNVHFQ 714
```

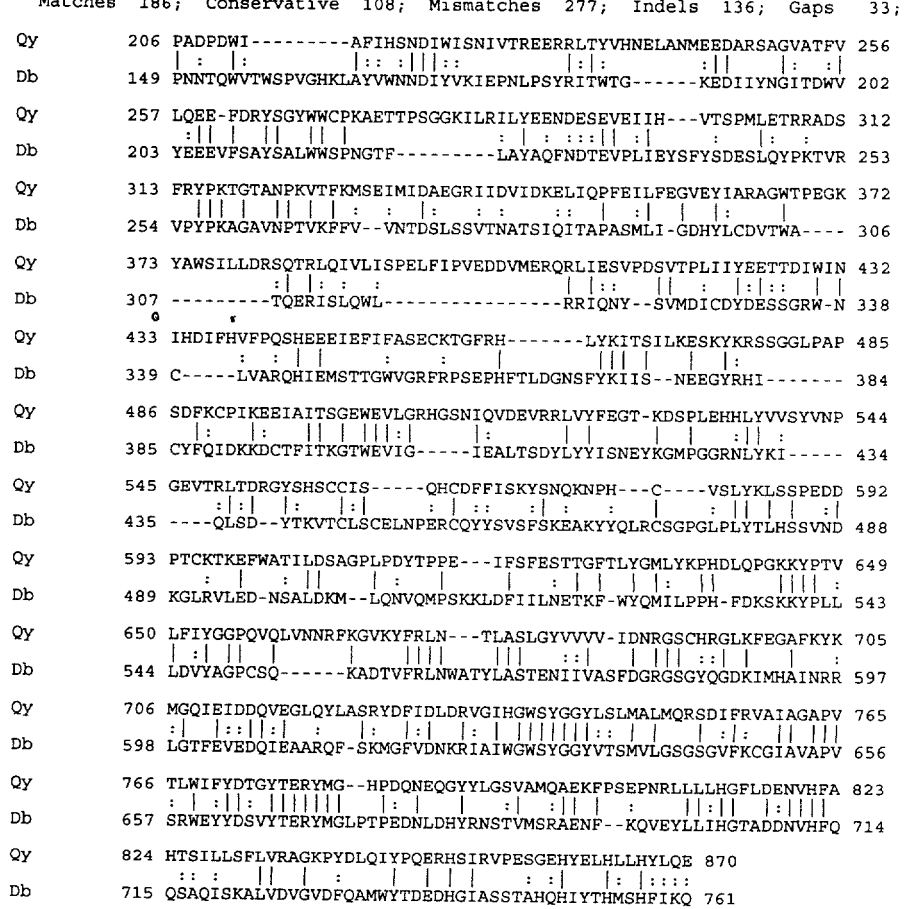
```
Qy        824 HTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQE 870
              :: :    ||  |   :     |  | | |     : :|    |: |::::
Db        715 QSAQISKALVDVGVDFQAMWYTDEDHGIASSTAHQHIYTHMSHFIKQ 761

RESULT 7
US-08-619-280A-3
; Sequence 3, Application US/08619280A
; Patent No. 5767242
;  GENERAL INFORMATION:
;    APPLICANT:  Zimmermann, Rainer; Park, John E.;
;    APPLICANT:  Rettig, Wolfgang; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
;    TITLE OF INVENTION:  ALPHA, AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 2.0 MB  storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/619,280A
;      FILING DATE:  18-MARCH-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/230,491
;      FILING DATE:  20-APRIL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5767242man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5330.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  766 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-619-280A-3

  Query Match            11.2%;  Score 525;  DB 1;  Length 766;
  Best Local Similarity   26.3%;  Pred. No. 5.5e-44;
  Matches  186;  Conservative  108;  Mismatches  277;  Indels  136;  Gaps   33;

Qy        206 PADPDWI---------AFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFV 256
              | :  |:         |:: :|||::        |:|:          :||    |:  :|
Db        149 PNNTQWVTWSPVGHKLAYVWNNDIYVKIEPNLPSYRITWTG------KEDIIYNGITDWV 202

Qy        257 LQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH---VTSPMLETRRADS 312
                :||  |  ||  || |                 : |  : :::|| :|     :   |:  :
Db        203 YEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSFYSDESLQYPKTVR 253

Qy        313 FRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGK 372
                 ||| |  || | | :  :  |:   :  :    ::   |  :|  |  |  |:    |
Db        254 VPYPKAGAVNPTVKFFV--VNTDSLSSVTNATSIQITAPASMLI-GDHYLCDVTWA---- 306

Qy        373 YAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWIN 432
                             :| |: :  :                 | |::   ||   :   |:|::  | |
Db        307 ---------TQERISLQWL----------------RRIQNY--SVMDICDYDESSGRW-N 338

Qy        433 IHDIFHVFPQSHEEEIEFIFASECKTGFRH-------LYKITSILKESKYKRSSGGLPAP 485
                        :  : |  |       :      :      |          |||  |    |  |:
Db        339 C-----LVARQHIEMSTTGWVGRFRPSEPHFTLDGNSFYKIIS--NEEGYRHI------- 384

Qy        486 SDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGT-KDSPLEHHLYVVSYVNP 544
                  |:   |:    || | |||:|      |:        | |      | |   :|| :
Db        385 CYFQIDKKDCTFITKGTWEVIG-----IEALTSDYLYYISNEYKGMPGGRNLYKI----- 434

Qy        545 GEVTRLTDRGYSHSCCIS-----QHCDFFISKYSNQKNPH---C----VSLYKLSSPEDD 592
                  :|:|   |:    |:|         :  |  ::   :  :|  :       :     |     :  ||  |  |  :|
Db        435 ----QLSD--YTKVTCLSCELNPERCQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVND 488

Qy        593 PTCKTKEFWATILDSAGPLPDYTPPE---IFSFESTTGFTLYGMLYKPHDLQPGKKYPTV 649
                  :   |    :  ||      |  :      |        |     :  |  |    |  |  :    ||          |||| :
Db        489 KGLRVLED-NSALDKM--LQNVQMPSKKLDFIILNETKF-WYQMILPPH-FDKSKKYPLL 543

Qy        650 LFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLKFEGAFKYK 705
                  | :| ||   |        |       ||||     |||     ::|     |  |||  ::|  |      |       :
Db        544 LDVYAGPCSQ------KADTVFRLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRR 597

Qy        706 MGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPV 765
                  :|   |::||:|   :     |:    |:|    |:  |  |||||||::  |   |      |  :|:  || |||
Db        598 LGTFEVEDQIEAARQF-SKMGFVDNKRIAIWGWSYGGYVTSMVLGSGSGVFKCGIAVAPV 656

Qy        766 TLWIFYDTGYTERYMG--HPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFA 823
                  :  |  :||: ||||||     |:    |        |     :|    :||    |    :          ||:||    |:||||
Db        657 SRWEYYDSVYTERYMGLPTPEDNLDHYRNSTVMSRAENF--KQVEYLLIHGTADDNVHFQ 714

Qy        824 HTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQE 870
                  :: :    ||  |   :     |  | | |     : :|    |: |::::
Db        715 QSAQISKALVDVGVDFQAMWYTDEDHGIASSTAHQHIYTHMSHFIKQ 761

RESULT 8
US-08-940-391-3
; Sequence 3, Application US/08940391
; Patent No. 5965373
;  GENERAL INFORMATION:
;    APPLICANT:  Zimmermann, Rainer; Park, John E.;
;    APPLICANT:  Rettig, Wolfgang; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED DIMERIC FIBROBLAST ACTIVATION
;    TITLE OF INVENTION:  PROTEIN ALPHA, AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 2.0 MB  storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/940,391
;      FILING DATE:  01-OCT-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/619,280
;      FILING DATE:  18-MARCH-1996
;      APPLICATION NUMBER:  08/230,491
;      FILING DATE:  20-APRIL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5965373man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5330.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  766 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-940-391-3
```

Query Match 11.2%; Score 525; DB 2; Length 766;
Best Local Similarity 26.3%; Pred. No. 5.5e-44;
Matches 186; Conservative 108; Mismatches 277; Indels 136; Gaps 33;



RESULT 9
US-09-794-236-1
; Sequence 1, Application US/09794236
; Patent No. 6337069
; GENERAL INFORMATION:
; APPLICANT: Grouzmann, Eric
; APPLICANT: Lacroix, Jean-Silvain
; APPLICANT: Monod, Michel
; TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
; FILE REFERENCE: 81985/276823
; CURRENT APPLICATION NUMBER: US/09/794,236
; CURRENT FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 766
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-794-236-1

Query Match 11.2%; Score 525; DB 4; Length 766;
Best Local Similarity 26.3%; Pred. No. 5.5e-44;
Matches 186; Conservative 108; Mismatches 277; Indels 136; Gaps 33;



RESULT 10
US-08-230-491A-2
; Sequence 2, Application US/08230491A
; Patent No. 5587299
; GENERAL INFORMATION:
; APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
; APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
; TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN ' AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FELFE & LYNCH
; STREET: 805 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE

```
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  WORDPERFECT - ASC II
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/230,491A
;      FILING DATE:  20-APRIL-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5587299man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 330
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  760 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-230-491A-2

  Query Match             9.7%;  Score 454;  DB 1;  Length 760;
  Best Local Similarity   24.1%;  Pred. No. 9.5e-37;
  Matches  169;  Conservative  113;  Mismatches  269;  Indels  150;  Gaps   27;

Qy        212 IAFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEF--DRYSGYWW 269
              :|::: |:|::          ::|:       |  |:    |:  :| :||    :|:  ||
Db        162 LAYVYQNNIYLKQRPGDPPFQITF------NGRENKIFNGIPDWVYEEEMLPTKYA-LWW 214

Qy        270 CPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLET-RRADSFRYPKTGTANPKVTFK 328
              |        || |   | | :: :: :|  :    |   |  :  ||| |  || |
Db        215 SP-------NGKFL--AYAEFNDKDIPVIAYSYYGDEQYPRTINIPYPKAGAKNPVVRI- 264

Qy        329 MSEIMIDAEGRIIDVIDKELIQPFEI-----LFEGVEYIARAGWTPEGKYAWSILLDRSQ 383
                        |||      : | |:     :     | :   |  : :     |
Db        265 ----------FIIDTTYPAYVGPQEVPVPAMIASSDYYFSWLTWVTDERVCLQWL----- 309

Qy        384 TRLQIVLISPELFIPVEDDVMERQRLIESVPDSVT----------PLIIYEETTDIWINI 433
              |:| | :        :   :  :  | : :| |          |:  |:      |:
Db        310 KRVQNVSVLSICDFREDWQTWDCPKTQEHIEESRTGWAGGFFVSRPVFSYDA-----ISY 364

Qy        434 HDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIK 493
              : ||                  | |::|:: |                      |  ::
Db        365 YKIF-----------------SDKDGYKHIHYI---------------------KDTVE 385

Qy        494 EEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDR 553
              | ||||:|| :     ||          |  |:||          |     : |::
Db        386 NAIQITSGKWEAI-----NI----------FRVTQDSLFYSSNEFEEYPGRRNIYRISIG 430

Qy        554 GYSHS-CCISQH-----CDFFISKYSNQKNPHCVSLY----KLSSPEDDPTCKTKEFWAT 603
              |  |  |:: |     | :: : :|    : :  |     :|:  |  | :
Db        431 SYPPSKKCVTCHLRKERCQYYTASFSDYAKYYALVCYGPGIPISTLHDGRTDQE----IK 486

Qy        604 ILDSAGPLPD-----YTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQV 658
              ||:    | :       | |          ||:  :  |      |||| :: ::||||
Db        487 ILEENKELENALKNIQLPKEEIKKLEVDEITLWYKMILPPQFDRSKKYPLLIQVYGGPCS 546

Qy        659 QLVNNRFKGVKYFRLNTLASL----GYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQ 714
              | | :     | :| :: |    | |: ::| ||:  :| |   |    |:|  |::||
Db        547 QSVRS------VFAVNWISYLASKEGMVIALVDGRGTAFQGDKLLYAVYRKLGVYEVEDQ 600

Qy        715 VEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTG 774
              :  ::       |||  |: | ||||||||:| :||   : :|:  || |||: | :| :
Db        601 ITAVRKFI-EMGFIDEKRIAIWGWSYGGYVSSLALASGTGLFKCGIAVAPVSSWEYYASV 659

Qy        775 YTERYMGHP--DQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFL 832
              ||||:|| |  | | : |   :|  :|| | :     ||:||  |:|||| ::: :   |
Db        660 YTERFMGLPTKDDNLEHYKNSTVMARAEYFRNVD--YLLIHGTADDNVHFQNSAQIAKAL 717

Qy        833 VRAGKPYDLQIYPQERHSIRVPESG---EHYELHLLHYLQE 870
              | |   :   | : | :    ||    |   |: |:|::
Db        718 VNAQVDFQAMWYSDQNHGL----SGLSTNHLYTHMTHFLKQ 754
```

```
RESULT 11
US-08-619-280A-2
; Sequence 2, Application US/08619280A
; Patent No. 5767242
;  GENERAL INFORMATION:
;    APPLICANT:  Zimmermann, Rainer; Park, John E.;
;    APPLICANT:  Rettig, Wolfgang; Old, Lloyd J.
;    TITLE OF INVENTION:  ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
;    TITLE OF INVENTION:  ALPHA, AND USES THEREOF
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 2.0 MB  storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/619,280A
;      FILING DATE:  18-MARCH-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/230,491
;      FILING DATE:  20-APRIL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5767242man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5330.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  760 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-619-280A-2

  Query Match             9.7%;  Score 454;  DB 1;  Length 760;
  Best Local Similarity   24.1%;  Pred. No. 9.5e-37;
  Matches  169;  Conservative  113;  Mismatches  269;  Indels  150;  Gaps   27;

Qy        212 IAFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEF--DRYSGYWW 269
              :|::: |:|::          ::|:       |  |:    |:  :| :||    :|:  ||
Db        162 LAYVYQNNIYLKQRPGDPPFQITF------NGRENKIFNGIPDWVYEEEMLPTKYA-LWW 214

Qy        270 CPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLET-RRADSFRYPKTGTANPKVTFK 328
              |        || |   | | :: :: :|  :    |   |  :  ||| |  || |
Db        215 SP-------NGKFL--AYAEFNDKDIPVIAYSYYGDEQYPRTINIPYPKAGAKNPVVRI- 264

Qy        329 MSEIMIDAEGRIIDVIDKELIQPFEI-----LFEGVEYIARAGWTPEGKYAWSILLDRSQ 383
                        |||      : | |:     :     | :   |  : :     |
Db        265 ----------FIIDTTYPAYVGPQEVPVPAMIASSDYYFSWLTWVTDERVCLQWL----- 309

Qy        384 TRLQIVLISPELFIPVEDDVMERQRLIESVPDSVT----------PLIIYEETTDIWINI 433
              |:| | :        :   :  :  | : :| |          |:  |:      |:
Db        310 KRVQNVSVLSICDFREDWQTWDCPKTQEHIEESRTGWAGGFFVSRPVFSYDA-----ISY 364

Qy        434 HDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIK 493
              : ||                  | |::|:: |                      |  ::
Db        365 YKIF-----------------SDKDGYKHIHYI---------------------KDTVE 385

Qy        494 EEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDR 553
              | ||||:|| :     ||          |  |:||          |     : |::
Db        386 NAIQITSGKWEAI-----NI----------FRVTQDSLFYSSNEFEEYPGRRNIYRISIG 430

Qy        554 GYSHS-CCISQH-----CDFFISKYSNQKNPHCVSLY----KLSSPEDDPTCKTKEFWAT 603
```

```
Db        431 SYPPSKKCVTCHLRKERCQYYTASFSDYAKYYALVCYGPGIPISTLHDGRTDQE----IK 486

Qy        604 ILDSAGPLPD-----YTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQV 658
Db        487 ILEENKELENALKNIQLPKEEIKKLEVDEITLWYKMILPPQFDRSKKYPLLIQVYGGPCS 546

Qy        659 QLVNNRFKGVKYFRLNTLASL----GYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQ 714
Db        547 QSVRS------VFAVNWISYLASKEGMVIALVDGRGTAFQGDKLLYAVYRKLGVYEVEDQ 600

Qy        715 VEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTG 774
Db        601 ITAVRKFI-EMGFIDEKRIAIWGWSYGGYVSSLALASGTGLFKCGIAVAPVSSWEYYASV 659

Qy        775 YTERYMGHP--DQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFL 832
Db        660 YTERFMGLPTKDDNLEHYKNSTVMARAEYFRNVD--YLLIHGTADDNVHFQNSAQIAKAL 717

Qy        833 VRAGKPYDLQIYPQERHSIRVPESG---EHYELHLLHYLQE 870
Db        718 VNAQVDFQAMWYSDQNHGL----SGLSTNHLYTHMTHFLKQ 754
```

RESULT 12
US-08-940-391-2
; Sequence 2, Application US/08940391
; Patent No. 5965373
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATIÓN
; TITLE OF INVENTION: PROTEIN ALPHA, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940,391
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/619,280
; FILING DATE: 18-MARCH-1996
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5965373man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 760 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-940-391-2

Query Match 9.7%; Score 454; DB 2; Length 760;
Best Local Similarity 24.1%; Pred. No. 9.5e-37;
Matches 169; Conservative 113; Mismatches 269; Indels 150; Gaps 27;

```
Qy        212 IAFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEF--DRYSGYWW 269
Db        162 LAYVYQNNIYLKQRPGDPPFQITF------NGRENKIFNGIPDWVYEEEMLPTKYA-LWW 214

Qy        270 CPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLET-RRADSFRYPKTGTANPKVTFK 328
Db        215 SP-------NGKFL--AYAEFNDKDIPVIAYSYYGDEQYPRTINIPYPKAGAKNPVVRI- 264

Qy        329 MSEIMIDAEGRIIDVIDKELIQPPEI-----LFEGVEYIARAGWTPEGKYAWSILLDRSQ 383
Db        265 ----------FIIDTTYPAYVGPQEVPVPAMIASSDYYFSWLTWVTDERVCLQWL----- 309

Qy        384 TRLQIVLISPELFIPVEDDVMERQRLIESVPDSVT----------PLIIYEETTDIWINI 433
Db        310 KRVQNVSVLSICDFREDWQTWDCPKTQEHIEESRTGWAGGFFVSRPVFSYDA-----ISY 364

Qy        434 HDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIK 493
Db        365 YKIF----------------SDKDGYKHIHYI----------------------KDTVE 385

Qy        494 EEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDR 553
Db        386 NAIQITSGKWEAI-----NI----------FRVTQDSLFYSSNEFEEYPGRRNIYRISIG 430

Qy        554 GYSHS-CCISQH-----CDFFISKYSNQKNPHCVSLY----KLSSPEDDPTCKTKEFWAT 603
Db        431 SYPPSKKCVTCHLRKERCQYYTASFSDYAKYYALVCYGPGIPISTLHDGRTDQE----IK 486

Qy        604 ILDSAGPLPD-----YTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQV 658
Db        487 ILEENKELENALKNIQLPKEEIKKLEVDEITLWYKMILPPQFDRSKKYPLLIQVYGGPCS 546

Qy        659 QLVNNRFKGVKYFRLNTLASL----GYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQ 714
Db        547 QSVRS------VFAVNWISYLASKEGMVIALVDGRGTAFQGDKLLYAVYRKLGVYEVEDQ 600

Qy        715 VEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTG 774
Db        601 ITAVRKFI-EMGFIDEKRIAIWGWSYGGYVSSLALASGTGLFKCGIAVAPVSSWEYYASV 659

Qy        775 YTERYMGHP--DQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFL 832
Db        660 YTERFMGLPTKDDNLEHYKNSTVMARAEYFRNVD--YLLIHGTADDNVHFQNSAQIAKAL 717

Qy        833 VRAGKPYDLQIYPQERHSIRVPESG---EHYELHLLHYLQE 870
Db        718 VNAQVDFQAMWYSDQNHGL----SGLSTNHLYTHMTHFLKQ 754
```

RESULT 13
US-09-016-080-1
; Sequence 1, Application US/09016080
; Patent No. 6133012
; GENERAL INFORMATION:
; APPLICANT: Ishikawa, Kaxuhiko
; APPLICANT: Matsui, Ikuo
; APPLICANT: Ishida, Hiroyasu
; APPLICANT: Kosugi, Yoshisugu
; APPLICANT: Higuchi, Katsuhiko
; TITLE OF INVENTION: THERMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 07898/022001
; CURRENT APPLICATION NUMBER: US/09/016,080
; CURRENT FILING DATE: 1998-01-30
; EARLIER APPLICATION NUMBER: JAPAN 18381/1997
; EARLIER FILING DATE: 1997-01-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 632
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-09-016-080-1

```
  Query Match             6.0%;  Score 282;  DB 3;  Length 632;
  Best Local Similarity   23.9%;  Pred. No. 2.4e-19;
  Matches  151;  Conservative   91;  Mismatches  242;  Indels  148;  Gaps   33;

Qy        276 TPSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIM-- 333
Db         67 SPNGKLIAFTSKRDKEGKESELYVIPTDGGEARLLAKFKY---GIKNLRFTEDGKSIAVV 123

Qy        334 --IDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLD---RSQTRL-- 386
Db        124 TPIDVEKKGND--DVHIIREIPFWFNGV------GWI-YGKRNVVYLVDVESGKKKRLTP 174

Qy        387 ------QIVLISPELFIPVEDD-----------VMERQ----------RLIESVPDSVTP 419
Db        175 KNLNVDQIRFHNGRLYFTAQEDRERKPLISDLYVLENRKVRKLTPGKWRILDFLPLDDGS 234

Qy        420 LIIYEETTD--IWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKR 477
Db        235 FVLKANTLERGIPTNAH-IYHYDP--------------KTG--ELKKLTKDLDRNAYNS 276

Qy        478 SSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLY 537
Db        277 LN------SDVRGSQRAELVYKEG-W-----------------IYYVAT-DGP-RANLF 309

Qy        538 VVSYVNPGEVTRLT--DRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTC 595
Db        310 RVNL--DGKIERVIGGDRSVE---------SFDIGDYIAFTAQDAVTPTELYIYRDGKEK 358

Qy        596 KTKEF--WATILDSAGPLPDYT--PPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLF 651
Db        359 KVTDFNKW---------IKGYTLSKPEHFKVKASDGVEIDAWVMKPVNFRKGKKYPAILE 409

Qy        652 IYGGPQVQLVNNRFKGVKYF-RLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIE 710
Db        410 IHGGPKTAY------GYAFMHEFHVLTSKGFVVIFSNPRGSDGYGEEF-ADIRGHYGERD 462

Qy        711 IDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIF 770
Db        463 YQDLMEVVDEALRRFDFIDGERLGVTGGSYGGFMTNW-IVGHTNRFKAAVTQRSISNWIS 521

Qy        771 Y----DTGYTERYMGHPDQ-------NEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDEN 819
Db        522 FFGTTDIGY---YFA-PDQIGKDPWSNLEGYWEKSPLKYA---PNVETPLLIIHSTEDYR 574

Qy        820 VHFAHTSILLSFLVRAGKPYDLQIYPQERHSI 851
Db        575 CWLPEALQLFISLKYLGKRVELAIFPGENHDL 606


RESULT 14
US-09-355-166-1
; Sequence 1, Application US/09355166
; Patent No. 6316241
; GENERAL INFORMATION:
;  APPLICANT:  Genencor International, Inc.
;  TITLE OF INVENTION:  Alpha/Beta Hydrolase-Fold Enzymes
;  FILE REFERENCE:  GC511-PCT
;  CURRENT APPLICATION NUMBER: US/09/355,166
;  CURRENT FILING DATE:  1999-07-20
;  NUMBER OF SEQ ID NOS:  21
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 657
;   TYPE: PRT
;   ORGANISM: Bacillus
US-09-355-166-1

  Query Match             5.3%;  Score 249.5;  DB 4;  Length 657;
  Best Local Similarity   21.4%;  Pred. No. 5.3e-16;
  Matches  150;  Conservative  101;  Mismatches  265;  Indels  185;  Gaps   31;

Qy        245 EDARSAGVATFVLQEEFDRYSGYWWCPKAET-----------------TPSGGKILRIL 286
Db         24 DGTRAAYVKSQVNQEK-DSYTSNIWIYETKTGGSVPWTHGEKRSTDPRWSPDGRTLAFIS 82

Qy        287 YEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIMID---AEGRIIDV 343
Db         83 DREGDAAQLYIMSTEGG--EARKLTDIPY---GVSKPLWSPDGESILVTISLGEGESIDD 137

Qy        344 IDK---ELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVE 400
Db        138 REKTEQDSYEPVEV--QGLSYKRDGKGLTRGAYA-----------QLVLVSVK------ 177

Qy        401 DDVMERQRLIESVPDSVTPL-------IIYEET---TDIWINIHDIFHVFPQSHEEEIEF 450
Db        178 --SGEMKELTSHKADHGDPAFSPDGKWLVFSANLTETDDASKPHDVY----------IMS 225

Qy        451 IFASECK--TGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEW----- 503
Db        226 LESGDLKQVTPHRGSFGSSSFSPDGRYLALLGN---EKEYK------NATLSKAWLYDIE 276

Qy        504 --------EVLGRH------GSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTR 549
Db        277 QGRLTCLTEMLDVHLADALIGDSLIGGAEQRPIW---TKDS---QGFYVIG--------- 321

Qy        550 LTDRG--------------------YSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSS 588
Db        322 -TDQGSTGIYYISIEGLVYPIRLEKEYINSFSLSPDEQHFIASVTKPDRPSELYSIPLGQ 380

Qy        589 PEDDPTCKTKEFWATILDSAGPLPDYT--PPEIFSFESTTGFTLYGMLYKPHDLQPGKKY 646
Db        381 EEKQLTGANDKF----------VREHTISIPEEIQYATEDGVMVNGWLMRPAQMEGETTY 430

Qy        647 PTVLFIYGGPQVQLVNNRFKGVKYF-RLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYK 705
Db        431 PLILNIHGGPHMMY------GHTYFHEFQVLAAKGYAVVYINPRGSHGYGQEFVNAVRGD 484

Qy        706 MGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPV 765
Db        485 YGGKDYDDVMQAVDEAIKRDPHIDPKRLGVTGGSYGGFMTNW-IVGQTNRFKAAVTQRSI 543

Qy        766 TLWIFY----DTGY--TERYMGHP---------DQNEQGYYLGSVAMQAEKFPSEPNRLL 810
Db        544 SNWISFHGVSDIGYFFTDWQLEHDMFEDTEKLWDRSPLKY-----AANVE------TPLL 592

Qy        811 LLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSI 851
Db        593 ILHGERDDRCPIEQAEQLFIALKKMGKETKLVRFPNASHNL 633


RESULT 15
PCT-US93-07923-11
; Sequence 11, Application PC/TUS9307923
;  GENERAL INFORMATION:
;    APPLICANT:  Morimoto, Chikao
;    APPLICANT:  Schlossman, Stuart F.
;    APPLICANT:  Tanaka, Toshiaki
;    TITLE OF INVENTION:  HUMAN CD26 AND METHODS FOR USE
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;      SOFTWARE:  WordPerfect (Version 5.0)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/07923
;      FILING DATE:  19930819
;      CLASSIFICATION:
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/934,162
;      FILING DATE:  21-AUG-1992
;      APPLICATION NUMBER:  07/832,211
;      FILING DATE:  06-FEB-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fraser, Janis K.
;      REGISTRATION NUMBER:  34,819
;      REFERENCE/DOCKET NUMBER:  00530/055002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  593
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
PCT-US93-07923-11

  Query Match             4.8%;  Score 225.5;  DB 5;  Length 593;
  Best Local Similarity   23.1%;  Pred. No. 1.2e-13;
  Matches  123;  Conservative   75;  Mismatches  203;  Indels  131;  Gaps   30;

Qy        206 PADPDWI---------AFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFV 256
              |  :  |:         |:: :|||::         |:|:        :||    |:  :|
Db        149 PNNTQWVTWSPVGHKLAYVWNNDIYVKIEPNLPSYRITWTG------KEDIIYNGITDWV 202

Qy        257 LQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH---VTSPMLETRRADS 312
               :|| |  ||  || |              : | : :::|| :|     :   |:  :
Db        203 YEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSFYSDESLQYPKTVR 253

Qy        313 FRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGK 372
             |||  |  || | |  :  :  |:   : :    ::  |  :|  |  |:    |
Db        254 VPYPKAGAVNPTVKFFV--VNTDSLSSVTNATSIQITAPASMLI-GDHYLCDVTWA---- 306

Qy        373 YAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWIN 432
                  :| |: :  :                 | |::   ||  :  |:|::  | |
Db        307 ---------TQERISLQWL---------------RRIQNY--SVMDICDYDESSGRW-N 338

Qy        433 IHDIFHVFPQSHEEEIEFIFASECKTGFRH-------LYKITSILKESKYKRSSGGLPAP 485
                  :  : | |     :    :    |          ||| |   |  |:
Db        339 C-----LVARQHIEMSTTGWVGRFRPSEPHFTLDGNSFYKIIS--NEEGYRHI------- 384

Qy        486 SDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGT-KDSPLEHHLYVVSYVNP 544
              |:   |:   || | |||:|     |:      | |      |  |   :|| :
Db        385 CYFQIDKKDCTFITKGTWEVIG-----IEALTSDYLYYISNEYKGMPGGRNLYKI----- 434

Qy        545 GEVTRLTDRGYSHSCCIS-----QHCDFFISKYSNQKNPH---C----VSLYKLSSPEDD 592
                :|:|  |:   |:|      : | ::   :| :   :    |    : || | |  :|
Db        435 ----QLSD--YTKVTCLSCELNPERCQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVND 488

Qy        593 PTCKTKEFWATILDSAGPLPDYTPPE---IFSFESTTGFTLYGMLYKPHDLQPGKKYPTV 649
               :  |   : ||    | :   |      |   : | |  | |:  ||     ||||  :
Db        489 KGLRVLED-NSALDKM--LQNVQMPSKKLDFIILNETKF-WYQMILPPH-FDKSKKYPLL 543

Qy        650 LFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLK 697
              | :| ||  |      |   ||||    |||   ::|   | ||| ::| |
Db        544 LDVYAGPCSQ------KADTVFRLNWATYLASTENIIVASFDGRGSGYQGDK 589


Search completed: October 15, 2003, 17:15:34
Job time : 23.6777 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:13:57 ; Search time 124.626 Seconds
(without alignments)
1140.337 Million cell updates/sec

Title: US-10-070-464-1
Perfect score: 4700
Sequence: 1 MAAAMETEQLGVEIFETADC.....HLLHYLQENLGSRIAALKVI 882

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 600653 seqs, 161128416 residues

Total number of hits satisfying chosen parameters: 600653

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4700	100.0	882	10	US-09-976-674-1	Sequence 1, Appli
2	4700	100.0	882	12	US-10-054-776-2	Sequence 2, Appli
3	4700	100.0	882	12	US-10-170-789-38	Sequence 38, Appl
4	3504	74.6	658	10	US-09-976-674-19	Sequence 19, Appl
5	3504	74.6	661	10	US-09-976-674-11	Sequence 11, Appl
6	3504	74.6	690	10	US-09-976-674-7	Sequence 7, Appli
7	3236	68.9	613	10	US-09-976-674-21	Sequence 21, Appl
8	2870	61.1	863	10	US-09-976-674-3	Sequence 3, Appli
9	2870	61.1	892	10	US-09-976-674-23	Sequence 23, Appl
10	2870	61.1	892	10	US-09-976-674-27	Sequence 27, Appl
11	2820.5	60.0	879	10	US-09-976-674-33	Sequence 33, Appl
12	2820.5	60.0	879	10	US-09-976-674-35	Sequence 35, Appl
13	2406	51.2	832	10	US-09-976-674-29	Sequence 29, Appl
14	2406	51.2	832	10	US-09-976-674-31	Sequence 31, Appl
15	2356.5	50.1	819	10	US-09-976-674-37	Sequence 37, Appl
16	2356.5	50.1	819	10	US-09-976-674-39	Sequence 39, Appl
17	1808	38.5	358	10	US-09-976-674-13	Sequence 13, Appl
18	1645.5	35.0	310	12	US-09-993-959-4	Sequence 4, Appli
19	1643.5	35.0	508	15	US-10-237-271-3	Sequence 3, Appli
20	1391	29.6	518	10	US-09-976-674-25	Sequence 25, Appl
21	1278	27.2	241	10	US-09-976-674-9	Sequence 9, Appli
22	1049	22.3	432	15	US-10-237-271-4	Sequence 4, Appli
23	1007.5	21.4	194	10	US-09-976-674-17	Sequence 17, Appl
24	585	12.4	129	11	US-09-764-891-3564	Sequence 3564, Ap
25	529	11.3	766	12	US-10-423-714-6	Sequence 6, Appli
26	529	11.3	766	14	US-10-002-593-6	Sequence 6, Appli
27	529	11.3	766	15	US-10-165-603-7	Sequence 7, Appli
28	525	11.2	766	9	US-09-265-606-3	Sequence 3, Appli
29	525	11.2	766	12	US-09-993-959-1	Sequence 1, Appli
30	523.5	11.1	767	15	US-10-165-603-4	Sequence 4, Appli
31	495	10.5	710	15	US-10-156-761-10681	Sequence 10681, A
32	484.5	10.3	818	12	US-10-401-437-3	Sequence 3, Appli
33	484.5	10.3	818	12	US-10-402-312-3	Sequence 3, Appli
34	484.5	10.3	818	12	US-10-402-067-3	Sequence 3, Appli
35	484.5	10.3	818	12	US-10-401-436-3	Sequence 3, Appli
36	464	9.9	108	10	US-09-976-674-15	Sequence 15, Appl
37	461	9.8	706	10	US-09-976-674-41	Sequence 41, Appl
38	461	9.8	796	10	US-09-976-674-5	Sequence 5, Appli
39	461	9.8	796	12	US-09-870-133-2	Sequence 2, Appli
40	461	9.8	796	15	US-10-160-501-5	Sequence 5, Appli
41	454	9.7	760	9	US-09-265-606-2	Sequence 2, Appli
42	454	9.7	760	12	US-10-301-822-55	Sequence 55, Appl
43	454	9.7	760	15	US-10-177-293-136	Sequence 136, App
44	431	9.2	803	12	US-10-205-219-58	Sequence 58, Appl
45	425.5	9.1	691	10	US-09-976-674-43	Sequence 43, Appl

ALIGNMENTS

RESULT 1
US-09-976-674-1
; Sequence 1, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-1

Query Match 100.0%; Score 4700; DB 10; Length 882;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
```

```
Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660

Qy        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720

Qy        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780

Qy        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840

Qy        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
              |||||||||||||||||||||||||||||||||||||||||||
Db        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
```

RESULT 2
US-10-054-776-2
; Sequence 2, Application US/10054776
; Publication No. US20030165818A1
; GENERAL INFORMATION:
; APPLICANT: Mark Robert Edbrooke
; APPLICANT: Alan Peter Lewis
; TITLE OF INVENTION: NOVEL PROTEIN
; FILE REFERENCE: QG1042US
; CURRENT APPLICATION NUMBER: US/10/054,776
; CURRENT FILING DATE: 2002-01-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-054-776-2

Query Match 100.0%; Score 4700; DB 12; Length 882;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660

Qy        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720

Qy        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780

Qy        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840

Qy        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
              |||||||||||||||||||||||||||||||||||||||||||
Db        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
```

RESULT 3
US-10-170-789-38
; Sequence 38, Application US/10170789
; Publication No. US20030180930A1
; GENERAL INFORMATION:
; APPLICANT: Rachel E. Meyers
; APPLICANT: Olandt, Peter J.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Curtis , Rory A. J.
; APPLICANT: Williamson, Mark
; APPLICANT: Weich, Nadine
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,

```
;  TITLE OF INVENTION:  AND PROTEASE FAMILY MEMBERS AND USES THEREOF
;  FILE REFERENCE: 10448-191001
;  CURRENT APPLICATION NUMBER: US/10/170,789
;  CURRENT FILING DATE:  2002-06-13
;  PRIOR APPLICATION NUMBER: US 09/797,039
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: PCT/US01/06525
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: US 60/186,061
;  PRIOR FILING DATE: 2000-02-29
;  PRIOR APPLICATION NUMBER: US 09/882,166
;  PRIOR FILING DATE: 2001-06-15
;  PRIOR APPLICATION NUMBER: PCT/US01/19269
;  PRIOR FILING DATE: 2001-06-15
;  PRIOR APPLICATION NUMBER: US 60/212,078
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 09/934,406
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: PCT/US01/26052
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: US 60/226,740
;  PRIOR FILING DATE: 2000-08-21
;  PRIOR APPLICATION NUMBER: US 09/861,801
;  PRIOR FILING DATE: 2001-05-21
;  PRIOR APPLICATION NUMBER: PCT/US01/16549
;  PRIOR FILING DATE: 2001-05-21
;  PRIOR APPLICATION NUMBER: US 60/205,508
;  PRIOR FILING DATE: 2000-05-19
;  PRIOR APPLICATION NUMBER: US 09/801,267
;  PRIOR FILING DATE: 2001-03-06
;  PRIOR APPLICATION NUMBER: PCT/US01/07138
;  PRIOR FILING DATE: 2001-03-05
;  PRIOR APPLICATION NUMBER: US 60/187,454
;  PRIOR FILING DATE: 2000-03-07
;  PRIOR APPLICATION NUMBER: US 09/829,671
;  PRIOR FILING DATE: 2001-04-10
;  PRIOR APPLICATION NUMBER: PCT/US01/40483
;  PRIOR FILING DATE: 2001-04-11
;  PRIOR APPLICATION NUMBER: US 60/197,508
;  PRIOR FILING DATE: 2000-04-18
;  PRIOR APPLICATION NUMBER: US 09/961,721
;  PRIOR FILING DATE: 2001-09-24
;  PRIOR APPLICATION NUMBER: PCT/US01/29904
;  PRIOR FILING DATE: 2001-09-24
;  PRIOR APPLICATION NUMBER: US 60/235,023
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US 10/045,367
;  PRIOR FILING DATE: 2001-11-07
;  PRIOR APPLICATION NUMBER: US 60/246,561
;  PRIOR FILING DATE: 2000-11-07
;  PRIOR APPLICATION NUMBER: US 09/801,275
;  PRIOR FILING DATE: 2001-03-06
;  PRIOR APPLICATION NUMBER: PCT/US01/07074
;  PRIOR FILING DATE: 2001-03-05
;  PRIOR APPLICATION NUMBER: US 60/187,420
;  PRIOR FILING DATE: 2000-03-07
;  NUMBER OF SEQ ID NOS: 63
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 38
;   LENGTH: 882
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-170-789-38

  Query Match             100.0%;  Score 4700;  DB 12;  Length 882;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  882;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWININHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWININHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660

Qy        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720

Qy        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780

Qy        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840

Qy        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
              |||||||||||||||||||||||||||||||||||||||||||
Db        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882


RESULT 4
US-09-976-674-19
; Sequence 19, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
```

; SEQ ID NO 19
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-19

Query Match 74.6%; Score 3504; DB 10; Length 658;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 655
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 655
```

RESULT 5
US-09-976-674-11
; Sequence 11, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 661
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-11

Query Match 74.6%; Score 3504; DB 10; Length 661;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 655
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 655
```

RESULT 6
US-09-976-674-7
; Sequence 7, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 690

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-7

Query Match 74.6%; Score 3504; DB 10; Length 690;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWININHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWININHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 655
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 655
```

RESULT 7
US-09-976-674-21
; Sequence 21, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-21

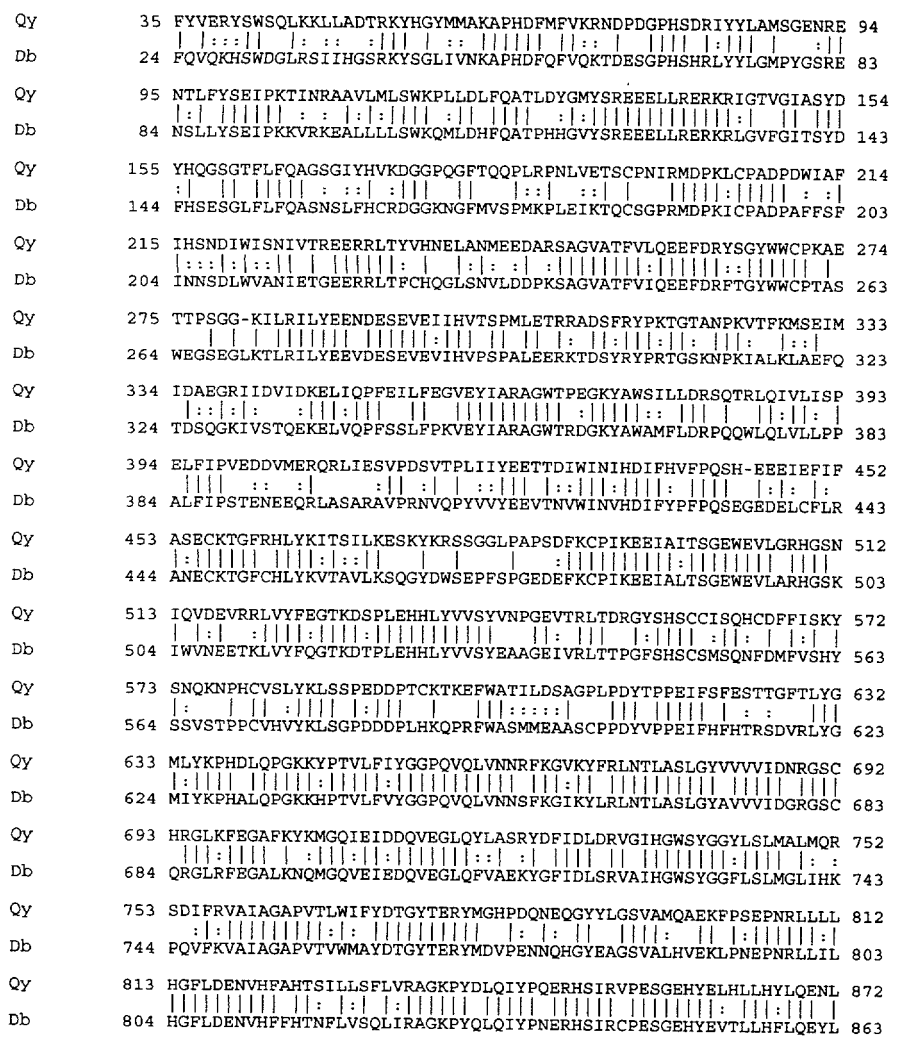
Query Match 68.9%; Score 3236; DB 10; Length 613;
Best Local Similarity 100.0%; Pred. No. 1.4e-301;
Matches 607; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWININHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWININHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDS 607
              |||||||
Db        601 WATILDS 607
```

RESULT 8
US-09-976-674-3
; Sequence 3, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 863
; TYPE: PRT
; ORGANISM: Homo sapiens

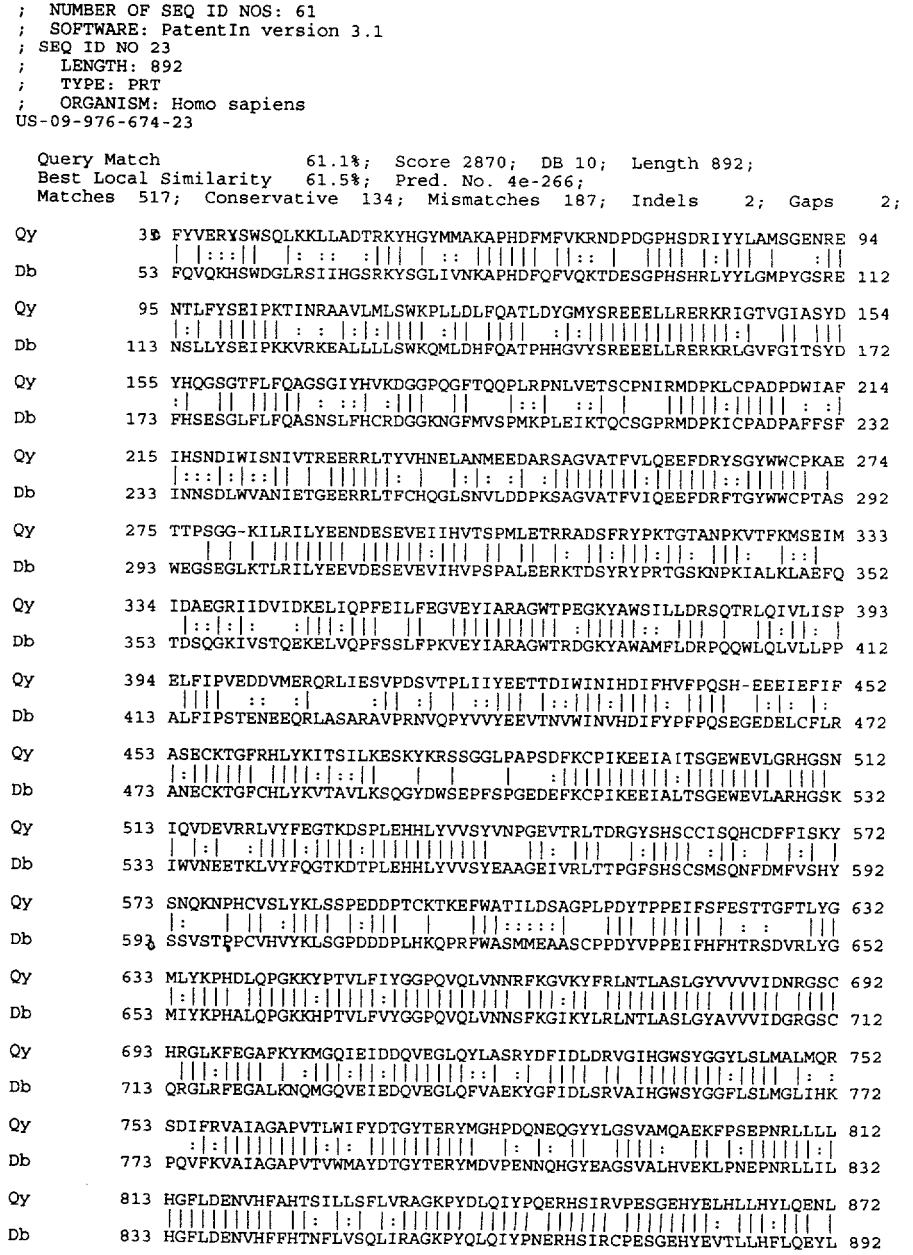
US-09-976-674-3

Query Match 61.1%; Score 2870; DB 10; Length 863;
Best Local Similarity 61.5%; Pred. No. 3.7e-266;
Matches 517; Conservative 134; Mismatches 187; Indels 2; Gaps 2;



RESULT 9
US-09-976-674-23
; Sequence 23, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-23

Query Match 61.1%; Score 2870; DB 10; Length 892;
Best Local Similarity 61.5%; Pred. No. 4e-266;
Matches 517; Conservative 134; Mismatches 187; Indels 2; Gaps 2;



RESULT 10
US-09-976-674-27
; Sequence 27, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-27

Query Match 61.1%; Score 2870; DB 10; Length 892;
Best Local Similarity 61.5%; Pred. No. 4e-266;
Matches 517; Conservative 134; Mismatches 187; Indels 2; Gaps 2;

```
Qy   35 FYVERYSWSQLKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRE 94
        | |:::|| |: :: :||| | :: |||||| ||:: | |||| |:||| | :||
Db   53 FQVQKHSWDGLRSIIHGSRKYSGLIVNKAPHDFQFVQKTDESGPHSHRLYYLGMPYGSRE 112

Qy   95 NTLFYSEIPKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYD 154
        |:| |||||| : : |:|:|||| :|| |||| :|:|||||||||||||:| || |||
Db  113 NSLLYSEIPKKVRKEALLLLSWKQMLDHFQATPHHGVYSREEELLRERKRLGVFGITSYD 172

Qy  155 YHQGSGTFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAF 214
        :| || ||||| : ::| :||| || |::| ::| | |||||:||||| : :|
Db  173 FHSESGLFLFQASNSLFHCRDGGKNGFMVSPMKPLEIKTQCSGPRMDPKICPADPAFFSF 232

Qy  215 IHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAE 274
        |:::|:|::|| | ||||||: | |:|: :| :||||||||:||||||::|||||| |
Db  233 INNSDLWVANIETGEERRLTFCHQGLSNVLDDPKSAGVATFVIQEEFDRFTGYWWCPTAS 292

Qy  275 TTPSGG-KILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIM 333
        | | | ||||||| ||||||:||| || || |: ||:|||:||: |||: |::|
Db  293 WEGSEGLKTLRILYEEVDESEVEVIHVPSPALEERKTDSYRYPRTGSKNPKIALKLAEFQ 352

Qy  334 IDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLISP 393
        |::|:|: :|||:||| || |||||||||| :|||||:: ||| | ||:||: |
Db  353 TDSQGKIVSTQEKELVQPFSSLFPKVEYIARAGWTRDGKYAWAMFLDRPQQWLQLVLLPP 412

Qy  394 ELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSH-EEEIEFIF 452
        |||| :: :| :|| :| | ::||| |::|||:|||||: |||| |:|: |:
Db  413 ALFIPSTENEEQRLASARAVPRNVQPYVVYEEVTNVWINVHDIFYPFPQSEGEDELCFLR 472

Qy  453 ASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSN 512
        |:|||||| ||||:|::|| | | | :|||||||||:|||||||| ||||
Db  473 ANECKTGFCHLYKVTAVLKSQGYDWSEPFSPGEDEFKCPIKEEIALTSGEWEVLARHGSK 532

Qy  513 IQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKY 572
        | |:| :||||:||||:|||||||||||  ||: ||| |:|||| :||: | |:| |
Db  533 IWVNEETKLVYFQGTKDTPLEHHLYVVSYEAAGEIVRLTTPGFSHSCSMSQNFDMFVSHY 592

Qy  573 SNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYG 632
        |: | || :|||| |:||| | |||:::::| ||| ||||| | : : |||
Db  593 SSVSTPPCVHVYKLSGPDDDPLHKQPRFWASMMEAASCPPDYVPPEIFHFHTRSDVRLYG 652

Qy  633 MLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSC 692
        |:|||| ||||||:|||||:|||||||||| |||:|| |||||||||| ||||| ||||
Db  653 MIYKPHALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNTLASLGYAVVVIDGRGSC 712

Qy  693 HRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQR 752
        |||:|||| | :|||:||:|||||||::| :| |||| || |||||||:|||| |: :
Db  713 QRGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAIHGWSYGGFLSLMGLIHK 772

Qy  753 SDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLL 812
        :|:|||||||||:|: |||||||||| |: |: || |||| || |:||||:|
Db  773 PQVFKVAIAGAPVTVWMAYDTGYTERYMDVPENNQHGYEAGSVALHVEKLPNEPNRLLIL 832

Qy  813 HGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQENL 872
        |||||||||| ||: |:| |:|||||| ||||| |||||| |||||||: |||:||| |
Db  833 HGFLDENVHFFHTNFLVSQLIRAGKPYQLQIYPNERHSIRCPESGEHYEVTLLHFLQEYL 892
```

RESULT 11
US-09-976-674-33
; Sequence 33, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-33

Query Match 60.0%; Score 2820.5; DB 10; Length 879;
Best Local Similarity 60.7%; Pred. No. 2.2e-261;
Matches 510; Conservative 132; Mismatches 183; Indels 15; Gaps 3;

```
Qy   35 FYVERYSWSQLKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRE 94
        | |:::|| |: :: :||| | :: |||||| ||:: | |||| |:||| | :||
Db   53 FQVQKHSWDGLRSIIHGSRKYSGLIVNKAPHDFQFVQKTDESGPHSHRLYYLGMPYGSRE 112

Qy   95 NTLFYSEIPKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYD 154
        |:| |||||| : : |:|:|||| :|| |||| :|:|||||||||||||:| || |||
Db  113 NSLLYSEIPKKVRKEALLLLSWKQMLDHFQATPHHGVYSREEELLRERKRLGVFGITSYD 172

Qy  155 YHQGSGTFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAF 214
        :| || ||||| : ::| :||| || |::| ::| | |||||:||||| : :|
Db  173 FHSESGLFLFQASNSLFHCRDGGKNGFMVSPMKPLEIKTQCSGPRMDPKICPADPAFFSF 232

Qy  215 IHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAE 274
        |:::|:|::|| | ||||||: | |:|: :| :||||||||:||||||::|||||| |
Db  233 INNSDLWVANIETGEERRLTFCHQGLSNVLDDPKSAGVATFVIQEEFDRFTGYWWCPTAS 292

Qy  275 TTPSGG-KILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIM 333
        | | | ||||||| ||||||:||| || || |: ||:|||:||: |||: |::|
Db  293 WEGSEGLKTLRILYEEVDESEVEVIHVPSPALEERKTDSYRYPRTGSKNPKIALKLAEFQ 352

Qy  334 IDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLISP 393
        |::|:|: :|||:||| || |||||||||| :|||||:: ||| | ||:||: |
Db  353 TDSQGKIVSTQEKELVQPFSSLFPKVEYIARAGWTRDGKYAWAMFLDRPQQWLQLVLLPP 412

Qy  394 ELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSH-EEEIEFIF 452
        |||| :: :| :|| :| | ::||| |::|||:|||||: |||| |:|: |:
Db  413 ALFIPSTENEEQRLASARAVPRNVQPYVVYEEVTNVWINVHDIFYPFPQSEGEDELCFLR 472

Qy  453 ASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSN 512
        |:|||||| ||||:|::|| | | | :|||||||||:|||||||| ||||
Db  473 ANECKTGFCHLYKVTAVLKSQGYDWSEPFSPGEDEFKCPIKEEIALTSGEWEVLARHGS- 531

Qy  513 IQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKY 572
                :||||:|||||||||||  ||: |||  |:|||| :||: | |:| |
```

```
Db        532 ------------KGTKDTPLEHHLYVVSYEAAGEIVRLTTPGFSHSCSMSQNFDMFVSHY 579

Qy        573 SNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYG 632

Db        580 SSVSTPPCVHVYKLSGPDDDPLHKQPRFWASMMEAASCPPDYVPPEIFHFHTRSDVRLYG 639

Qy        633 MLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSC 692

Db        640 MIYKPHALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNTLASLGYAVVVIDGRGSC 699

Qy        693 HRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQR 752

Db        700 QRGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAIHGWSYGGFLSLMGLIHK 759

Qy        753 SDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLL 812

Db        760 PQVFKVAIAGAPVTVWMAYDTGYTERYMDVPENNQHGYEAGSVALHVEKLPNEPNRLLIL 819

Qy        813 HGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQENL 872

Db        820 HGFLDENVHFFHTNFLVSQLIRAGKPYQLQIYPNERHSIRCPESGEHYEVTLLHFLQEYL 879
```

RESULT 12
US-09-976-674-35
; Sequence 35, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-35

Query Match 60.0%; Score 2820.5; DB 10; Length 879;
Best Local Similarity 60.7%; Pred. No. 2.2e-261;
Matches 510; Conservative 132; Mismatches 183; Indels 15; Gaps 3;

```
Qy         35 FYVERYSWSQLKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRE 94

Db         53 FQVQKHSWDGLRSIIHGSRKYSGLIVNKAPHDFQFVQKTDESGPHSHRLYYLGMPYGSRE 112

Qy         95 NTLFYSEIPKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYD 154

Db        113 NSLLYSEIPKKVRKEALLLLSWKQMLDHFQATPHHGVYSREEELLRERKRLGVFGITSYD 172

Qy        155 YHQGSGTFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAF 214

Db        173 FHSESGLFLFQASNSLFHCRDGGKNGFMVSPMKPLEIKTQCSGPRMDPKICPADPAFFSF 232

Qy        215 IHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAE 274

Db        233 INNSDLWVANIETGEERRLTFCHQGLSNVLDDPKSAGVATFVIQEEFDRFTGYWWCPTAS 292

Qy        275 TTPSGG-KILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIM 333

Db        293 WEGSEGLKTLRILYEEVDESEVEVIHVPSPALEERKTDSYRYPRTGSKNPKIALKLAEFQ 352

Qy        334 IDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLISP 393

Db        353 TDSQGKIVSTQEKELVQPFSSLFPKVEYIARAGWTRDGKYAWAMFLDRPQQWLQLVLLPP 412
```

```
Qy        394 ELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSH-EEEIEFIF 452

Db        413 ALFIPSTENEEQRLASARAVPRNVQPYVVYEEVTNVWINVHDIFYPFPQSEGEDELCFLR 472

Qy        453 ASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSN 512

Db        473 ANECKTGFCHLYKVTAVLKSQGYDWSEPFSPGEDEFKCPIKEEIALTSGEWEVLARHGS- 531

Qy        513 IQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKY 572

Db        532 ------------KGTKDTPLEHHLYVVSYEAAGEIVRLTTPGFSHSCSMSQNFDMFVSHY 579

Qy        573 SNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYG 632

Db        580 SSVSTPPCVHVYKLSGPDDDPLHKQPRFWASMMEAASCPPDYVPPEIFHFHTRSDVRLYG 639

Qy        633 MLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSC 692

Db        640 MIYKPHALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNTLASLGYAVVVIDGRGSC 699

Qy        693 HRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQR 752

Db        700 QRGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAIHGWSYGGFLSLMGLIHK 759

Qy        753 SDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLL 812

Db        760 PQVFKVAIAGAPVTVWMAYDTGYTERYMDVPENNQHGYEAGSVALHVEKLPNEPNRLLIL 819

Qy        813 HGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQENL 872

Db        820 HGFLDENVHFFHTNFLVSQLIRAGKPYQLQIYPNERHSIRCPESGEHYEVTLLHFLQEYL 879
```

RESULT 13
US-09-976-674-29
; Sequence 29, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976,674
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-674-29

Query Match 51.2%; Score 2406; DB 10; Length 832;
Best Local Similarity 59.0%; Pred. No. 1.4e-221;
Matches 434; Conservative 122; Mismatches 177; Indels 2; Gaps 2;

```
Qy         35 FYVERYSWSQLKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRE 94

Db         53 FQVQKHSWDGLRSIIHGSRKYSGLIVNKAPHDFQFVQKTDESGPHSHRLYYLGMPYGSRE 112

Qy         95 NTLFYSEIPKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYD 154

Db        113 NSLLYSEIPKKVRKEALLLLSWKQMLDHFQATPHHGVYSREEELLRERKRLGVFGITSYD 172

Qy        155 YHQGSGTFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAF 214

Db        173 FHSESGLFLFQASNSLFHCRDGGKNGFMVSPMKPLEIKTQCSGPRMDPKICPADPAFFSF 232
```

```
Qy        215 IHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAE 274
              |:::|:|::|| | |||||:  |  |:|: :|  :|||||||:||||||::|||||| |
Db        233 INNSDLWVANIETGEERRLTFCHQGLSNVLDDPKSAGVATFVIQEEFDRFTGYWWCPTAS 292

Qy        275 TTPSGG-KILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIM 333
                | | | ||||||| ||||||:||| || || |: ||:|||:||: |||:  |::|
Db        293 WEGSEGLKTLRILYEEVDESEVEVIHVPSPALEERKTDSYRYPRTGSKNPKIALKLAEFQ 352

Qy        334 IDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLISP 393
              |::|:|:   :|||:|||  ||  |||||||||| :|||||:: ||| |  ||:||: |
Db        353 TDSQGKIVSTQEKELVQPFSSLFPKVEYIARAGWTRDGKYAWAMFLDRPQQWLQLVLLPP 412

Qy        394 ELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSH-EEEIEFIF 452
              ||||  ::  :|    :|| :| | ::||| |::|||:||||: |||  |:|: |:
Db        413 ALFIPSTENEEQRLASARAVPRNVQPYVVYEEVTNVWINVHDIFYPFPQSEGEDELCFLR 472

Qy        453 ASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSN 512
              |:|||||| ||||:|::||   |  |    |  :||||||||:|||||||| ||||
Db        473 ANECKTGFCHLYKVTAVLKSQGYDWSEPFSPGEDEFKCPIKEEIALTSGEWEVLARHGSK 532

Qy        513 IQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKY 572
              | |:|  :||||:||||:||||||||||   ||: |||  |:|||| :||: | |:| |
Db        533 IWVNEETKLVYFQGTKDTPLEHHLYVVSYEAAGEIVRLTTPGFSHSCSMSQNFDMFVSHY 592

Qy        573 SNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYG 632
              |:   | || :|||| |:|||  |   |||:::::|   ||| |||||  | : :   |||
Db        593 SSVSTPPCVHVYKLSGPDDDPLHKQPRFWASMMEAASCPPDYVPPEIFHFHTRSDVRLYG 652

Qy        633 MLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSC 692
              |:|||| |||||:|||||:|||||||||| |||:|| |||||||||| ||||| ||||
Db        653 MIYKPHALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNTLASLGYAVVVIDGRGSC 712

Qy        693 HRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQR 752
              |||:|||| | :|||:||:|||||||::| :| |||| || ||||||||:|||| |: :
Db        713 QRGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAIHGWSYGGFLSLMGLIHK 772

Qy        753 SDIFRVAIAGAPVTL 767
              :|:     |  |
Db        773 PQVFKAQPLAYPPRL 787


RESULT 14
US-09-976-674-31
; Sequence 31, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
;   LENGTH: 832
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-976-674-31

  Query Match            51.2%;  Score 2406;  DB 10;  Length 832;
  Best Local Similarity  59.0%;  Pred. No. 1.4e-221;
  Matches  434;  Conservative  122;  Mismatches  177;  Indels    2;  Gaps    2;

Qy         35 FYVERYSWSQLKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRE 94
              | |:::||  |: ::  :||| | :: |||||| ||:: |  |||| |:||| |   :||
Db         53 FQVQKHSWDGLRSIIHGSRKYSGLIVNKAPHDFQFVQKTDESGPHSHRLYYLGMPYGSRE 112

Qy         95 NTLFYSEIPKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYD 154
                |:| |||||| : : |:|:|||| :|| ||||  :|:|||||||||||||:|  || |||
Db        113 NSLLYSEIPKKVRKEALLLLSWKQMLDHFQATPHHGVYSREEELLRERKRLGVFGITSYD 172

Qy        155 YHQGSGTFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAF 214
              :|  || ||||| : ::| :|||  ||   |::|  ::| |   |||||:||||| : :|
Db        173 FHSESGLFLFQASNSLFHCRDGGKNGFMVSPMKPLEIKTQCSGPRMDPKICPADPAFFSF 232

Qy        215 IHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAE 274
              |:::|:|::|| | |||||:  |  |:|: :|  :|||||||:||||||::|||||| |
Db        233 INNSDLWVANIETGEERRLTFCHQGLSNVLDDPKSAGVATFVIQEEFDRFTGYWWCPTAS 292

Qy        275 TTPSGG-KILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIM 333
                | | | ||||||| ||||||:||| || || |: ||:|||:||: |||:  |::|
Db        293 WEGSEGLKTLRILYEEVDESEVEVIHVPSPALEERKTDSYRYPRTGSKNPKIALKLAEFQ 352

Qy        334 IDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLISP 393
              |::|:|:   :|||:|||  ||  |||||||||| :|||||:: ||| |  ||:||: |
Db        353 TDSQGKIVSTQEKELVQPFSSLFPKVEYIARAGWTRDGKYAWAMFLDRPQQWLQLVLLPP 412

Qy        394 ELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSH-EEEIEFIF 452
              ||||  ::  :|    :|| :| | ::||| |::|||:||||: |||  |:|: |:
Db        413 ALFIPSTENEEQRLASARAVPRNVQPYVVYEEVTNVWINVHDIFYPFPQSEGEDELCFLR 472

Qy        453 ASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSN 512
              |:|||||| ||||:|::||   |  |    |  :||||||||:|||||||| ||||
Db        473 ANECKTGFCHLYKVTAVLKSQGYDWSEPFSPGEDEFKCPIKEEIALTSGEWEVLARHGSK 532

Qy        513 IQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKY 572
              | |:|  :||||:||||:||||||||||   ||: |||  |:|||| :||: | |:| |
Db        533 IWVNEETKLVYFQGTKDTPLEHHLYVVSYEAAGEIVRLTTPGFSHSCSMSQNFDMFVSHY 592

Qy        573 SNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYG 632
              |:   | || :|||| |:|||  |   |||:::::|   ||| |||||  | : :   |||
Db        593 SSVSTPPCVHVYKLSGPDDDPLHKQPRFWASMMEAASCPPDYVPPEIFHFHTRSDVRLYG 652

Qy        633 MLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSC 692
              |:|||| |||||:|||||:|||||||||| |||:|| |||||||||| ||||| ||||
Db        653 MIYKPHALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNTLASLGYAVVVIDGRGSC 712

Qy        693 HRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQR 752
              |||:|||| | :|||:||:|||||||::| :| |||| || ||||||||:|||| |: :
Db        713 QRGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAIHGWSYGGFLSLMGLIHK 772

Qy        753 SDIFRVAIAGAPVTL 767
              :|:     |  |
Db        773 PQVFKAQPLAYPPRL 787


RESULT 15
US-09-976-674-37
; Sequence 37, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
;  APPLICANT: Qi, Steve
;  APPLICANT:  Akinsanya, Karen
;  APPLICANT:  Riviere, Pierre
;  APPLICANT:  Junien, Jean-Louis
;  TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
;  FILE REFERENCE: 70669
;  CURRENT APPLICATION NUMBER: US/09/976,674
;  CURRENT FILING DATE:  2001-10-12
;  PRIOR APPLICATION NUMBER: US 60/240,117
;  PRIOR FILING DATE: 2000-10-12
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37
;   LENGTH: 819
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-976-674-37

  Query Match            50.1%;  Score 2356.5;  DB 10;  Length 819;
  Best Local Similarity  58.1%;  Pred. No. 7.7e-217;
```

Matches 427; Conservative 120; Mismatches 173; Indels 15; Gaps 3;

```
Qy          35 FYVERYSWSQLKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRE 94
               | |:::||  |: ::  :||| | :: |||||| ||:: |  |||| |:||| |   :||
Db          53 FQVQKHSWDGLRSIIHGSRKYSGLIVNKAPHDFQFVQKTDESGPHSHRLYYLGMPYGSRE 112

Qy          95 NTLFYSEIPKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYD 154
               |:| |||||| : : |:|:|||| :|| |||| :|:|||||||||||||:|  || |||
Db         113 NSLLYSEIPKKVRKEALLLLSWKQMLDHFQATPHHGVYSREEELLRERKRLGVFGITSYD 172

Qy         155 YHQGSGTFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAF 214
               :|  || ||||| : ::| :||| ||   |::|  ::| |   |||||:||||| : :|
Db         173 FHSESGLFLFQASNSLFHCRDGGKNGFMVSPMKPLEIKTQCSGPRMDPKICPADPAFFSF 232

Qy         215 IHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAE 274
               |:::|:|::|| | ||||||: |  |:|: :| :|||||||:||||||::|||||| |
Db         233 INNSDLWVANIETGEERRLTFCHQGLSNVLDDPKSAGVATFVIQEEFDRFTGYWWCPTAS 292

Qy         275 TTPSGG-KILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIM 333
               | | | ||||||| ||||||:||| || || |: ||:|||:||: |||:  |::|
Db         293 WEGSEGLKTLRILYEEVDESEVEVIHVPSPALEERKTDSYRYPRTGSKNPKIALKLAEFQ 352

Qy         334 IDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLISP 393
               |::|:|:   :|||:||| || |||||||||| :|||||:: ||| |  ||:||: |
Db         353 TDSQGKIVSTQEKELVQPFSSLFPKVEYIARAGWTRDGKYAWAMFLDRPQQWLQLVLLPP 412

Qy         394 ELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSH-EEEIEFIF 452
               ||||  ::  :|    :|| :| | ::||| |::|||:||||: ||||  |:|: |:
Db         413 ALFIPSTENEEQRLASARAVPRNVQPYVVYEEVTNVWINVHDIFYPFPQSEGEDELCFLR 472

Qy         453 ASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSN 512
               |:|||||| ||||:|::||   | |    |   :|||||||||:|||||||| ||||
Db         473 ANECKTGFCHLYKVTAVLKSQGYDWSEPFSPGEDEFKCPIKEEIALTSGEWEVLARHGS- 531

Qy         513 IQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKY 572
                     :||||:||||||||||   ||: |||  |:|||| :||: | |:| |
Db         532 -----------KGTKDTPLEHHLYVVSYEAAGEIVRLTTPGFSHSCSMSQNFDMFVSHY 579

Qy         573 SNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYG 632
               |:   | || :|||| |:||  |   |||:::::|   ||| ||||| | : :   |||
Db         580 SSVSTPPCVHVYKLSGPDDDPLHKQPRFWASMMEAASCPPDYVPPEIFHFHTRSDVRLYG 639

Qy         633 MLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSC 692
               |:|||| |||||:|||||:|||||||||||| |||:|| |||||||||| ||||| ||||
Db         640 MIYKPHALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNTLASLGYAVVVIDGRGSC 699

Qy         693 HRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQR 752
               |||:|||| | :|||:||:|||||||::| :| |||| || |||||||:|||| |: :
Db         700 QRGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAIHGWSYGGFLSLMGLIHK 759

Qy         753 SDIFRVAIAGAPVTL 767
               :|:    |  |
Db         760 PQVFKAQPLAYPPRL 774
```

Search completed: October 15, 2003, 17:28:48
Job time : 126.626 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:08:55 ; Search time 23.176 Seconds
(without alignments)
3659.853 Million cell updates/sec

Title: US-10-070-464-1
Perfect score: 4700
Sequence: 1 MAAAMETEQLGVEIFETADC.....HLLHYLQENLGSRIAALKVI 882

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	922.5	19.6	931	2	T32919	hypothetical prote
2	842	17.9	738	2	A87516	dipeptidyl peptida
3	754	16.0	741	2	JC5142	X-Pro dipeptidyl-p
4	620	13.2	711	2	S66261	X-Pro dipeptidyl-p
5	529	11.3	766	1	CDHU26	dipeptidyl-peptida
6	519.5	11.1	793	2	T41703	dipeptidyl aminope
7	518.5	11.0	760	1	S23752	dipeptidyl-peptida
8	510.5	10.9	792	1	A39914	dipeptidyl-peptida
9	484.5	10.3	818	1	A30107	dipeptidyl aminope
10	465	9.9	931	2	A49737	dipeptidyl aminope
11	451.5	9.6	799	2	T25174	hypothetical prote
12	445.5	9.5	779	2	T25173	hypothetical prote
13	440	9.4	803	2	A41793	dipeptidyl aminope
14	427	9.1	803	2	I68600	dipeptidyl aminope
15	427	9.1	865	2	I54331	dipeptidyl aminope
16	378	8.0	829	2	T19514	hypothetical prote
17	374	8.0	795	2	F82858	dipeptidyl-peptida
18	331.5	7.1	759	2	I38593	fibroblast activat
19	293	6.2	743	2	T37700	probable dipeptidy
20	282	6.0	642	2	C71137	hypothetical prote
21	269.5	5.7	631	2	H75007	probable acylamino
22	268	5.7	683	2	E87495	prolyl oligopeptid
23	255.5	5.4	709	2	B82580	alanyl dipeptidyl
24	249.5	5.3	657	2	E70025	probable acylamino
25	236	5.0	622	2	F71174	hypothetical prote
26	223.5	4.8	632	2	E75057	peptidase PAB1418
27	217.5	4.6	536	2	F90299	acylaminoacyl-pept
28	200.5	4.3	591	2	H72474	probable acylamino
29	185	3.9	569	2	S74053	probable acylamino
30	183.5	3.9	608	2	F83397	probable peptidase
31	182.5	3.9	674	2	B84381	acylaminoacyl-pept
32	180.5	3.8	721	2	T09631	probable acylamino
33	180	3.8	667	2	A87711	prolyl oligopeptid
34	172	3.7	732	1	JC4655	acylaminoacyl-pept
35	171	3.6	732	1	S07624	acylaminoacyl-pept
36	170.5	3.6	572	2	F72455	probable acylamino
37	163.5	3.5	659	2	F72568	probable acylamino
38	161	3.4	676	2	C97775	acylamino-acid-rel
39	160.5	3.4	732	1	JU0132	acylaminoacyl-pept
40	160.5	3.4	745	2	T33751	hypothetical prote
41	158	3.4	598	2	F84199	hypothetical prote
42	157	3.3	614	2	E75094	prolyl endopeptida
43	155.5	3.3	654	2	AD3183	peptidase [importe
44	155	3.3	629	2	T15945	hypothetical prote
45	154.5	3.3	828	2	G87584	hypothetical prote

ALIGNMENTS

RESULT 1
T32919
hypothetical protein K02F2.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T32919
R;Maggi, L.; Goela, D.
submitted to the EMBL Data Library, January 1998
A;Description: The sequence of C. elegans cosmid K02F2.
A;Reference number: Z21246
A;Accession: T32919
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-931 <MAG>
A;Cross-references: EMBL:AF043699; PIDN:AAB97564.1; GSPDB:GN00019; CESP:K02F2.1
A;Experimental source: strain Bristol N2; clone K02F2
C;Genetics:
A;Gene: CESP:K02F2.1
A;Map position: 1
A;Introns: 58/3; 82/2; 131/2; 178/2; 275/3; 322/2; 404/2; 441/1; 464/2; 486/3; 528/2; 5

Query Match 19.6%; Score 922.5; DB 2; Length 931;
Best Local Similarity 29.3%; Pred. No. 1.4e-57;
Matches 282; Conservative 145; Mismatches 326; Indels 211; Gaps 41;

```
Qy         33 EPFYVERYSWSQLKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGEN 92
              ||   |  |:||   |:  | :  :       | :    :   ||:  |:|
Db         36 EPARFETRSFSQ---LIDHARSWKTEVRGMTTQGFTKISLMRAE---KDRLNMYAISSVP 89

Qy         93 RENT--LFYSEIP-KTINRAAVLMLSWKPLLDLFQATLDYGMY-------------SRE 135
              ||  :|   || : : :| |           |:  |  |                | |
Db         90 GTNTQSIFSVTIPLELVEKAQVADRK-------FELKLKSGYNVDSYIRMSCRKTPPSAE 142

Qy        136 EELLRERKRIGTV-GIASYDYHQG-----SGTFLFQAG---------------------- 167
               |  ||:|   | ||: |:   |     :|  ||:
Db        143 FTLQCERQRSQVVTGISDYEIRNGKMILMAGDQLFRYNPLNEALAAIPIAVPDDQSSTEP 202

Qy        168 ---------SGIYHVKDGGPQG-----FTQQPLRPNLVETSCP-------NIRMDPKLCP 206
                      ||      :  ||       |: |::     |  |       |      |:||
Db        203 MDISEGSITSGTKGCSNEAPQSSTVPPVTRIPIKKPTTSTEKPATAPPTNNFVSSAKVCP 262

Qy        207 ADPDWIAFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSG 266
              ||   :|:: :  ::|       |:    :|   :| :      || ::::||| :|: |
Db        263 ADSSLLAYVLNKQVYI-------EKNGKIIHRTSSNSKHITN--GVPSYIVQEELERFEG 313

Qy        267 YWWCPKAETTPSGGKILRILYEENDESEVEIIHV-------TSPMLETRRADSFRYPKTG 319
              ||  :::|        |:|||  :| :|            :||          :||: |
Db        314 IWW-SESKT--------RLLYEHVNEEKVAESQFGVNGDPPVAPM---------KYPRAG 355

Qy        320 TANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEG---VEYIARAGWTPEGKYAWS 376
              | |   | :|   :|   |:  ||  |:     |:::     ||| |||:  :|   |
Db        356 TKNAYSTLRM---VILENGKAYDVPLKD-----EVIYKHCPFYEYITRAGFFSDGTTVWV 407
```

```
Qy        377 ILLDRSQTRLQIVLI----------------------SPELFIPVEDDVMERQRLIESVP 414
Db        408 QVMSRDQAQCSLLLIPYTDFLLPEELGGSIKEDNLQLSTDLNMGVWDD-KSHEETMEKPP 466

Qy        415 DSVT--PLIIYEETTDIWININHDIFHVFPQSHEEE--IEFIFASECKTGFRHLYKITSIL 470
Db        467 RGKLRGTVQIHKARNDYWINTHNAIYPLKITDEEHPMYEFIYCLEKPNG-SCLALISAEL 525

Qy        471 KESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDS 530
Db        526 DQNGY--------------CRHTEEKLLMAENFSI--NKSMGIVVDEVRELVYYVANESH 569

Qy        531 PLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQK----NPHCVSLYKL 586
Db        570 PTEWNI-CVSHYRTGQHAQLTESGI---------C--FKSERANGKLALDLDHGFACYMT 617

Qy        587 S--SPEDDPTCKTKEF-W------------ATILDSAGP-LPD--YTPPEIFSFES-TTG 627
Db        618 SVGSPAE---CRFYSFRWKENEVLPSTVYAANITVSGHPGQPDLHFDSPEMIEFQSKKTG 674

Qy        628 FTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVID 687
Db        675 LMHYAMILRPSNFDPYKKYPVFHYVYGGPGIQIVHNDFSWIQYIR---FCRLGYVVVFID 731

Qy        688 NRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRY-DFIDLDRVGIHGWSYGGYLSL 746
Db        732 NRGSAHRGIEFERHIHKKMGTVEVEDQVEGLQMLAERTGGFMDMSRVVVHGWSYGGYMAL 791

Qy        747 MALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEP 806
Db        792 QMIAKHPNIYRAAIAGGAVSDWRLYDTAYTERYMGYP-LEEHVYGASSITGLVEKLPDEP 850

Qy        807 NRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLH 866
Db        851 NRLMLVHGLMDENVHFAHLTHLVDECIKKGKWHELVIFPNERHGVRNNDASIYLDARMMY 910

Qy        867 YLQE 870
Db        911 FAQQ 914
```

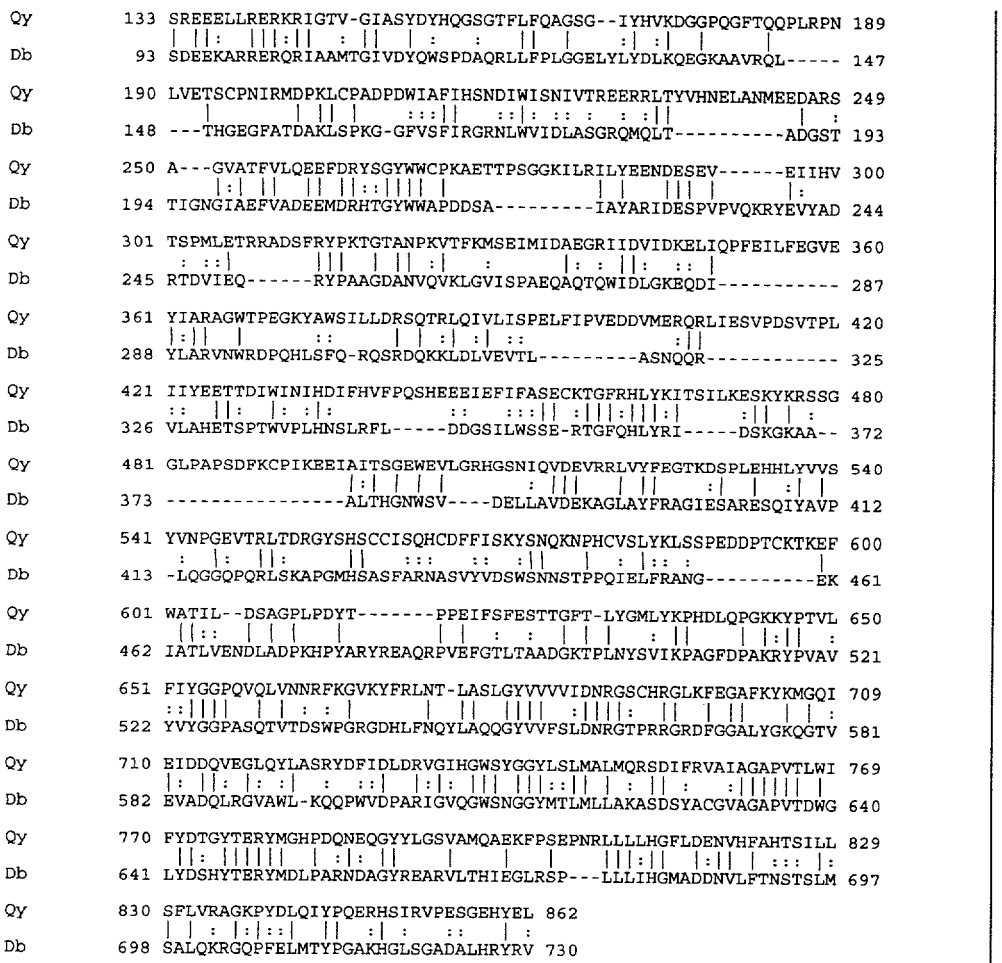
RESULT 2
A87516
dipeptidyl peptidase IV [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: A87516
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: A87516
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-738 <STO>
A;Cross-references: GB:AE005673; NID:g13423647; PIDN:AAK24125.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC2154

Query Match 17.9%; Score 842; DB 2; Length 738;
Best Local Similarity 29.9%; Pred. No. 5.5e-52;
Matches 221; Conservative 115; Mismatches 262; Indels 141; Gaps 24;

```
Qy        133 SREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIY--HVKDGGPQGFTQQPLRPNL 190
Db         94 SEAEKARRERARVSARGIVEYSWDRQGRFILVPLDGDLYLDAVADGKITRLTE------- 146

Qy        191 VETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSA 250
Db        147 ----TPGDEVDAKVSPKG-GYVSYVRDQNLYIKPVAGGAETALT-------TDGKDALSF 194

Qy        251 GVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLETRRA 310
Db        195 GVAEFIVQEELDRFTGYWWSPDES---------RIVYTRVDESGVDIV---------PRA 236

Qy        311 D---------SFRYPKTGTANPKVTFKMSEIMIDAEGRI--IDV-IDKELIQPFEILFEG 358
Db        237 DIGPGGATVVNQRYPRAGRPNAVVDLFVRDL---ASGKVTALDLGANKDI---------- 283

Qy        359 VEYIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVT 418
Db        284 --YVARVAWSADGKTVYVQRLSRDQKTLDLLAF----------------------DAAT 318

Qy        419 PL--IIYEETTDIWININHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYK 476
Db        319 GAGKTILTDTDPHFIEVSNDFRPLTDG-----TFLWGSE-KDGNQHLYRYA--------- 363

Qy        477 RSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHL 536
Db        364 -ADGKLIA-----------QITKGDWPVIGLEG----VDEARKVAIFSASIDTPIERRL 406

Qy        537 YVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSP-----ED 591
Db        407 YEVSYAKPGKPKALTSAGGWWAAKVADNGGAFAGTYSDPKTPSQTALYSADGKRVRWIEE 466

Qy        592 DPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLF 651
Db        467 NKLAEGHPYW----PYAANLPQ---PEFGSLKAADGETLHYEILKPIGFDPAKKYPAIVS 519

Qy        652 IYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEI 711
Db        520 VYGGPHAQRVMKNWHSPSE---RTYLEAGYVIFKLDNRGSGNRSAKFMRALDRKLGTVEV 576

Qy        712 DDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFY 771
Db        577 EDQLLGAKFLASQ-PYVDADKLGVMGWSYGGFMALMLLTAENTPFKAGAAGAPPTEWSLY 635

Qy        772 DTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSF 831
Db        636 DTAYTERYMGKPDENKAGYAYSDINNRIDKL--APGSLLLLHGMADDNVIFENSTRLMAA 693

Qy        832 LVRAGKPYDLQIYPQERHS 850
Db        694 LQRKAILFEMAMYPGERHS 712
```

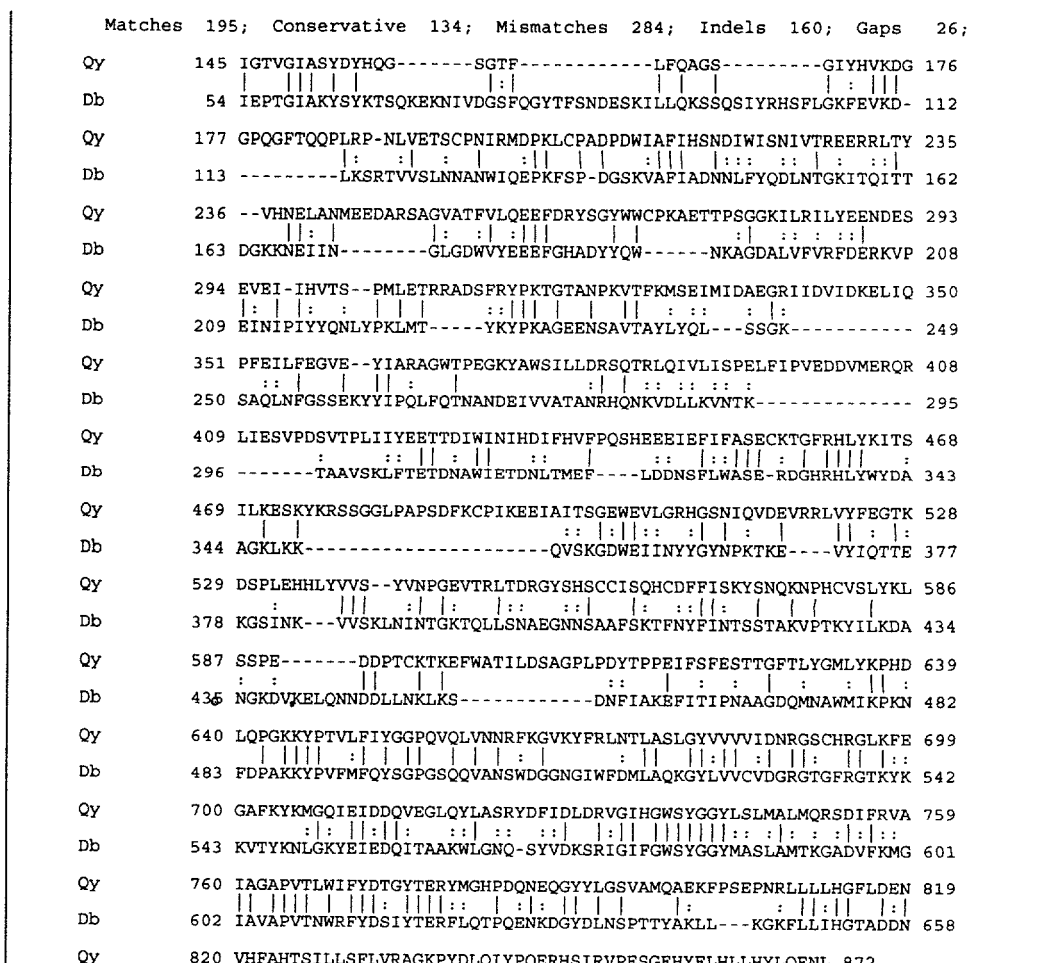
RESULT 3
JC5142
X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Xanthomonas maltophilia
C;Species: Xanthomonas maltophilia
C;Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 20-Jun-2000
C;Accession: JC5142
R;Kabashima, T.; Ito, K.; Yoshimoto, T.
J. Biochem. 120, 1111-1117, 1996
A;Title: Dipeptidyl peptidase IV from Xanthomonas maltophilia: Sequencing and expressio
A;Reference number: JC5142; MUID:97164011; PMID:9010758
A;Accession: JC5142
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-741 <KAB>
A;Cross-references: DDBJ:D83263; NID:g1753196; PIDN:BAA11872.1; PID:g1753197
C;Comment: This enzyme catalyzes the hydrolytic removal of amino-terminal dipeptidyl re
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptide hydrolase; membrane bound
F;4-18/Domain: transmembrane #status predicted <TMM>
F;610/Active site: Ser #status predicted
F;685/Active site: Asp #status predicted
F;717/Active site: His #status predicted

Query Match 16.0%; Score 754; DB 2; Length 741;
Best Local Similarity 28.4%; Pred. No. 1.1e-45;
Matches 214; Conservative 120; Mismatches 281; Indels 138; Gaps 26;



RESULT 4
S66261
X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Flavobacterium meningosepticum
C;Species: Flavobacterium meningosepticum
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C;Accession: S66261
R;Kabashima, T.; Yoshida, T.; Ito, K.; Yoshimoto, T.
Arch. Biochem. Biophys. 320, 123-128, 1995
A;Title: Cloning, sequencing, and expression of the dipeptidyl peptidase IV gene from Fl
A;Reference number: S66261; MUID:95314307; PMID:7793970
A;Accession: S66261
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-711 <KAB>
A;Cross-references: EMBL:D42121; NID:g577283; PIDN:BAA07702.1; PID:g577284
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptide hydrolase

Query Match 13.2%; Score 620; DB 2; Length 711;
Best Local Similarity 25.2%; Pred. No. 4e-36;



RESULT 5
CDHU26
dipeptidyl-peptidase IV (EC 3.4.14.5) - human
N;Alternate names: cell surface glycoprotein CD26; thymocyte-activating molecule (THAM)
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 23-Aug-1996 #text_change 18-Jun-1999
C;Accession: S24313; B42408; A42408; B61136; S59510; I56154; S59857; S15520
R;Misumi, Y.; Hayashi, Y.; Arakawa, F.; Ikehara, Y.
Biochim. Biophys. Acta 1131, 333-336, 1992
A;Title: Molecular cloning and sequence analysis of human dipeptidyl peptidase IV, a se
A;Reference number: S24313; MUID:92329551; PMID:1352704
A;Accession: S24313
A;Molecule type: mRNA
A;Residues: 1-6,'I',8-766 <MIS>
A;Cross-references: EMBL:X60708; NID:g35335; PIDN:CAA43118.1; PID:g35336
R;Darmoul, D.; Lacasa, M.; Baricault, L.; Marguet, D.; Sapin, C.; Trotot, P.; Barbat, A
J. Biol. Chem. 267, 4824-4833, 1992
A;Title: Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon cance

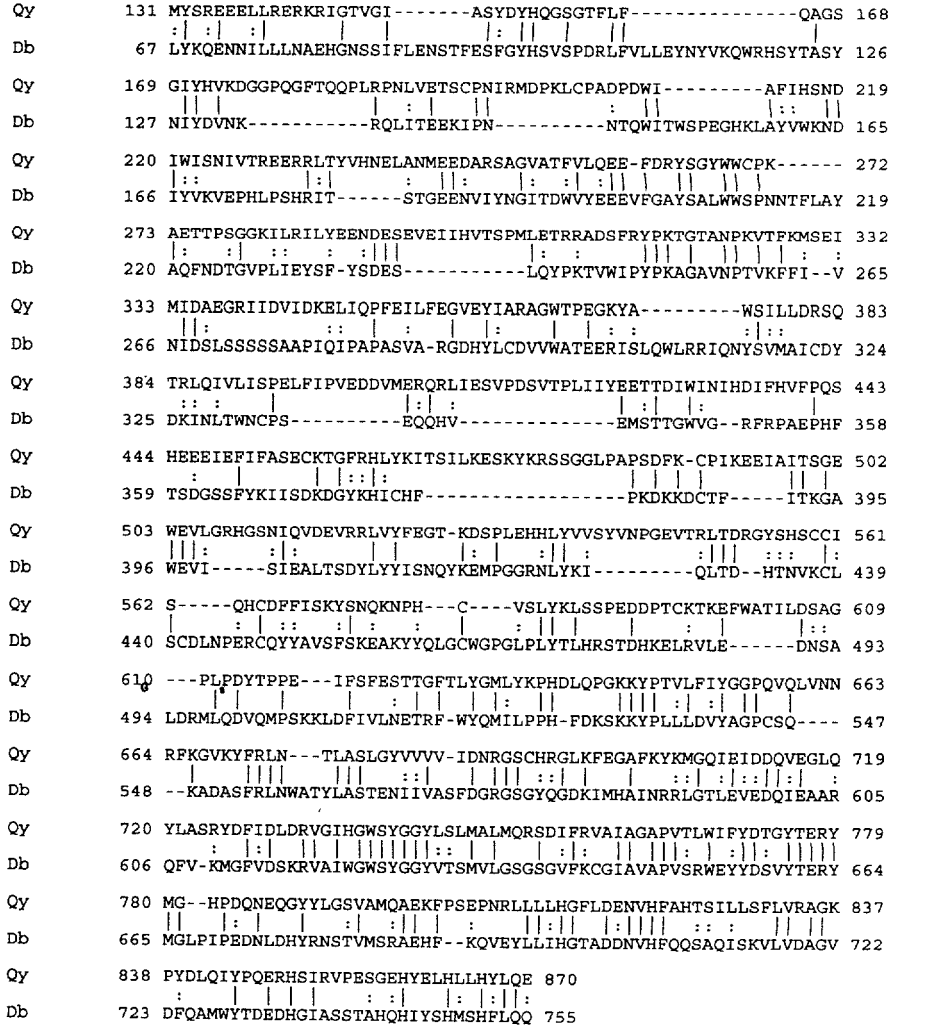
```
 IV mRNA levels during cell differentiation.
A;Reference number: A42408; MUID:92165847; PMID:1347043
A;Accession: B42408
A;Molecule type: mRNA
A;Residues: 1-5,'R',7-436,'S',438-556,'I',558-662,'E',664-766 <DAR1>
A;Cross-references: GB:M80536; NID:g181569; PIDN:AAA52308.1; PID:g181570
A;Experimental source: intestine
A;Note: this sequence corresponds with the author's translation
A;Accession: A42408
A;Molecule type: mRNA
A;Residues: 1-5,'R',7-436,'S',438-556,'I',558-662,'E',664-711,'G',713-766 <DAR2>
A;Cross-references: GB:M80536; NID:g181569
A;Note: sequence extracted from NCBI backbone (NCBIN:83986, NCBIP:83988); this sequence
R;Gorvel, J.P.; Ferrero, A.; Chambraud, L.; Rigal, A.; Bonicel, J.; Maroux, S.
Gastroenterology 101, 618-625, 1991
A;Title: Expression of sucrase-isomaltase and dipeptidylpeptidase IV in human small inte
A;Reference number: A61136; MUID:91317403; PMID:1677636
A;Accession: B61136
A;Molecule type: protein
A;Residues: 1-15,'X',17-22 <GOR>
R;Boehm, S.K.; Gum Jr., J.R.; Erickson, R.H.; Hicks, J.W.; Kim, Y.S.
Biochem. J. 311, 835-843, 1995
A;Title: Human dipeptidyl peptidase IV gene promoter: tissue-specific regulation from a
A;Reference number: S59510; MUID:96067599; PMID:7487939
A;Accession: S59510
A;Molecule type: DNA
A;Residues: 1-31 <BOE>
A;Cross-references: GB:S79876; NID:g1195574; PIDN:AAB35614.1; PID:g1195575
R;Tanaka, T.; Camerini, D.; Seed, B.; Torimoto, Y.; Dang, N.H.; Kameoka, J.; Dahlberg, H
J. Immunol. 149, 481-486, 1992
A;Title: Cloning and functional expression of the T cell activation antigen CD26.
A;Reference number: I56154; MUID:92325476; PMID:1352530
A;Accession: I56154
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-436,'S',438-766 <TAN>
A;Cross-references: GB:M74777; NID:g180082; PIDN:AAA51943.1; PID:g180083
R;Abbott, C.A.; Baker, E.; Sutherland, G.R.; McCaughan, G.W.
Immunogenetics 40, 331-338, 1994
A;Title: Genomic organization, exact localization, and tissue expression of the human CD
A;Reference number: S59857; MUID:95012454; PMID:7927537
A;Accession: S59857
A;Molecule type: DNA
A;Residues: 1-436,'S',438-766 <ABB>
A;Cross-references: EMBL:U13734
C;Genetics:
A;Gene: GDB:DPP4
A;Cross-references: GDB:125239; OMIM:102720
A;Map position: 2q24.3-2q24.3
A;Introns: 2/3; 32/1; 65/1; 95/3; 122/3; 140/2; 164/3; 205/1; 258/3; 296/2; 341/3; 356/3
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptide hydrolase; glycoprotein; homodimer; proteinase; transmembr
F;1-6/Domain: intracellular #status predicted <INT>
F;7-28/Domain: transmembrane #status predicted <TMN>
F;29-766/Domain: extracellular #status predicted <EXT>
F;85,92,150,219,229,281,321,520,685/Binding site: carbohydrate (Asn) (covalent) #status
F;630,708,740/Active site: Ser, Asp, His #status predicted

  Query Match             11.3%;  Score 529;  DB 1;  Length 766;
  Best Local Similarity   26.5%;  Pred. No. 1.4e-29;
  Matches  186;  Conservative  109;  Mismatches  280;  Indels  128;  Gaps   32;

Qy        206 PADPDWI---------AFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFV 256
               | :  |:          |: :||:        |:|:         :||     |:  :|
Db        149 PNNTQWVTWSPVGHKLAYVWNNDIYVKIEPNLPSYRITWTG------KEDIIYNGITDWV 202

Qy        257 LQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH---VTSPMLETRRADS 312
               :||  |  ||  ||  |               : |  : :::|| :|    :    |:  :
Db        203 YEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSFYSDESLQYPKTVR 253

Qy        313 FRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGK 372
                 |||  |   || | |  :   :   |:   : :     ::   |  :|   |  |:    |
Db        254 VPYPKAGAVNPTVKFFV--VNTDSLSSVTNATSIQITAPASMLI-GDHYLCDVTWA---- 306

Qy        373 YAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWIN 432
                       :| |: :  :                | |::    ||  :  |:|::  | |
Db        307 ---------TQERISLQWL---------------RRIQNY--SVMDICYDYESSGRW-N 338

Qy        433 IHDIFHVFPQSHEEEIEFIFASECKTGFRH-------LYKITSILKESKYKRSSGGLPAP 485
                :  : | |      :        :          |     |||  |    |  |:
Db        339 C-----LVARQHIEMSTTGWVGRFRPSEPHFTLDGNSFYKIIS--NEEGYRHI------- 384

Qy        486 SDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGT-KDSPLEHHLYVVSYVNP 544
                |:    |:    || |  |||:|        |:              |  |     |    :||  :  ::
Db        385 CYFQIDKKDCTFITKGTWEVIG-----IEALTSDYLYYISNEYKGMPGGRNLYKIQLIDY 439

Qy        545 GEVTRLTDRGYSHSCCIS-QHCDFFISKYSNQKNPH---C----VSLYKLSSPEDDPTCK 596
               :|| |         || :: :  | ::   :| :   :      |      : || | |   :|    :
Db        440 TKVTCL-------SCELNPERCQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVNDKGLR 492

Qy        597 TKEFWATILDSAGPLPDYTPPE---IFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIY 653
               |    : ||      | :     |        |    : |  |   | |:  ||       |||  :| :|
Db        493 VLED-NSALDKM--LQNVQMPSKKLDFIILNETKF-WYQMILPPH-FDKSKKYPLLLDVY 547

Qy        654 GGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLKFEGAFKYKMGQI 709
               ||   |         |   ||||         |||        ::|   | |||  ::|  |        ::|
Db        548 AGPCSQ------KADTVFRLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRRLGTF 601

Qy        710 EIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWI 769
               |::||:|      :     |:    |:|    |:  | |||||||:: |     |      :|:  || ||:  |
Db        602 EVEDQIEAARQF-SKMGFVDNKRIAIWGWSYGGYVTSMVLGSGSGVFKCGIAVAPVSRWE 660

Qy        770 FYDTGYTERYMG--HPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSI 827
               :||: ||||||     |:  |     |    :|  :|| |    :     ||:||   |:||||    ::
Db        661 YYDSVYTERYMGLPTPEDNLDHYRNSTVMSRAENF--KQVEYLLIHGTADDNVHFQQSAQ 718

Qy        828 LLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQE 870
               :   ||   |    :     |    |  |  |       :  :|     |: |::::
Db        719 ISKALVDVGVDFQAMWYTDEDHGIASSTAHQHIYTHMSHFIKQ 761
```

```
RESULT 6
T41703
dipeptidyl aminopeptidase - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C;Accession: T41703
R;Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1998
A;Reference number: Z22011
A;Accession: T41703
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-793 <MUR>
A;Cross-references: EMBL:AL031180; PIDN:CAA20138.1; GSPDB:GN00066
A;Experimental source: strain 972h-; cosmid c2E11 -chimeric
C;Genetics:
A;Gene: SPAC2E11.08
A;Map position: 1
C;Superfamily: dipeptidyl-peptidase IV

  Query Match             11.1%;  Score 519.5;  DB 2;  Length 793;
  Best Local Similarity   26.3%;  Pred. No. 7.3e-29;
  Matches  187;  Conservative  113;  Mismatches  256;  Indels  155;  Gaps   32;

Qy        212 IAFIHSNDIWI----SNIVTREERRLTY-----VHNELANMEEDARSAGVATFVLQEE-F 261
              ::|:::||:::      |:      :||||      | |                   |:  :: :||
Db        183 LSFVYNNDLYVRKNDGNV-----QRLTYDGTVDVFN------------GLTDWIYEEEVL 225

Qy        262 DRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV----------TSPMLETRRADS 312
                 |  || |  ::              :| : : :||:  |             :  |  :  :  :
Db        226 SSPSTIWWSPDSD---------KIAFLKLNESEIPTYHYPLYTAELDPSLPEFDYNKDMA 276

Qy        313 FRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGK 372
               :|||  |  ||  |:  ::::  :|              ::  |  :|     :  |  :  |:  :  :
Db        277 IKYPKPGNPNPSVSLFVADLNSNASSNFSLWHNEPLAEP---VVQNVLWVNTS------- 326
```

RESULT 7
S23752
dipeptidyl-peptidase IV (EC 3.4.14.5) alpha chain - mouse
N;Alternate names: CD26 alpha subunit; THAM alpha subunit
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S23752; A46465; A56030
R;Marguet, D.; Bernard, A.M.; Vivier, I.; Darmoul, D.; Naquet, P.; Pierres, M.
J. Biol. Chem. 267, 2200-2208, 1992
A;Title: cDNA cloning for mouse thymocyte-activating molecule. A multifunctional ecto-di
A;Reference number: S23752; MUID:92129288; PMID:1370813
A;Accession: S23752
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-760 <MAR>
A;Cross-references: EMBL:X58384
R;Vivier, I.; Marguet, D.; Naquet, P.; Bonicel, J.; Black, D.; Li, C.X.; Bernard, A.M.;
J. Immunol. 147, 447-454, 1991
A;Title: Evidence that thymocyte-activating molecule is mouse CD26 (dipeptidyl peptidase
A;Reference number: A46465; MUID:91302787; PMID:1712807
A;Accession: A46465
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-20 <VIV>
A;Experimental source: M14.T thymoma cells, Swiss nu/nu
A;Note: sequence extracted from NCBI backbone (NCBIP:42236)
R;Bernard, A.M.; Mattei, M.G.; Pierres, M.; Marguet, D.
Biochemistry 33, 15204-15214, 1994
A;Title: Structure of the mouse dipeptidyl peptidase IV (CD26) gene.
A;Reference number: A56030; MUID:95092780; PMID:7999781
A;Accession: A56030
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 746-760 <BER>
A;Cross-references: GB:U12620

C;Genetics:
A;Gene: CD26
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein
F;213,223,315,514,679/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;624,702,734/Active site: Ser, Asp, His #status predicted

Query Match 11.0%; Score 518.5; DB 1; Length 760;
Best Local Similarity 25.0%; Pred. No. 8e-29;
Matches 203; Conservative 107; Mismatches 306; Indels 197; Gaps 37;



RESULT 8
A39914
dipeptidyl-peptidase IV (EC 3.4.14.5), membrane-bound form precursor - rat
N;Alternate names: GP110; membrane glycoprotein 110K; OX-61
N;Contains: dipeptidyl-peptidase IV, soluble form

```
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A39914; A33315; B33315; A60730; A42203; S38949; A31781
R;Hong, W.; Doyle, D.
Proc. Natl. Acad. Sci. U.S.A. 84, 7962-7966, 1987
A;Title: cDNA cloning for a bile canaliculus domain-specific membrane glycoprotein of ra
A;Reference number: A39914; MUID:88068516; PMID:3479775
A;Accession: A39914
A;Molecule type: mRNA
A;Residues: 1-792 <HON>
A;Cross-references: GB:J02997; NID:g204463; PIDN:AAA41272.1; PID:g204464
R;Ogata, S.; Misumi, Y.; Ikehara, Y.
J. Biol. Chem. 264, 3596-3601, 1989
A;Title: Primary structure of rat liver dipeptidyl peptidase IV deduced from its cDNA an
A;Reference number: A33315; MUID:89123496; PMID:2563382
A;Accession: A33315
A;Molecule type: mRNA
A;Residues: 1-37,'A',39-182,'I',184-331,'T',333-351,'C',353-393,'V',395-561,'L',563-623,
A;Cross-references: GB:J04591; NID:g203973; PIDN:AAA41096.1; PID:g203974
A;Note: the authors translated the codon GCG for residue 38 as Arg, ACC for residue 332
A;Accession: B33315
A;Molecule type: protein
A;Residues: 1-20;35-54;427-443;505-509;511-520;530-538;593-600;602-608;618-627 <HO2>
R;McCaughan, G.W.; Wickson, J.E.; Creswick, P.F.; Gorrell, M.D.
Hepatology 11, 534-544, 1990
A;Title: Identification of the bile canalicular cell surface molecule GP110 as the ectop
quence.
A;Reference number: A60730; MUID:90228896; PMID:1970322
A;Accession: A60730
A;Molecule type: protein
A;Residues: 28-47,'XX',50-53,55-58 <MCC>
R;Ogata, S.; Misumi, Y.; Tsuji, E.; Takami, N.; Oda, K.; Ikehara, Y.
Biochemistry 31, 2582-2587, 1992
A;Title: Identification of the active site residues in dipeptidyl peptidase IV by affini
A;Reference number: A42203; MUID:92190188; PMID:1347701
A;Accession: A42203
A;Molecule type: protein
A;Residues: 'R',625-630,'X',632-648 <OG2>
R;Iwaki-Egawa, S.; Watanabe, Y.; Fujimoto, Y.
Biol. Chem. Hoppe-Seyler 374, 973-975, 1993
A;Title: N-terminal amino acid sequence of the 60-kDa protein of rat kidney dipeptidyl p
A;Reference number: S38949; MUID:94128239; PMID:7905271
A;Accession: S38949
A;Status: preliminary
A;Molecule type: protein
A;Residues: 281-302 <IWA>
R;Hong, W.; Doyle, D.
J. Biol. Chem. 263, 16892-16898, 1988
A;Title: Membrane orientation of rat gp110 as studied by in vitro translation.
A;Reference number: A31781; MUID:89034185; PMID:3182821
A;Accession: A31781
A;Molecule type: mRNA
A;Residues: 1-40 <HO3>
C;Comment: This protein is localized to the bile canaliculus, which is the apical domain
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptide hydrolase; glycoprotein; homodimer; liver; serine proteina
F;1-792/Product: dipeptidyl-peptidase, membrane-bound form #status experimental <MATM>
F;1-28/Domain: signal sequence #link MATS #status experimental <SIG>
F;1-6/Domain: intracellular #status predicted <INT>
F;7-28/Domain: transmembrane #status predicted <TMN>
F;29-792/Domain: extracellular #status predicted <EXT>
F;29-34/Domain: propeptide #link MATS #status experimental <PRO>
F;35-792/Product: dipeptidyl-peptidase, soluble form #status experimental <MATS>
F;83,90,148,217,227,319,521,686/Binding site: carbohydrate (Asn) (covalent) #status pred
F;631/Active site: Ser #status experimental
F;709,741/Active site: Asp, His #status predicted

  Query Match            10.9%;  Score 510.5;  DB 1;  Length 792;
  Best Local Similarity  25.7%;  Pred. No. 3.2e-28;
  Matches  182;  Conservative  105;  Mismatches  287;  Indels  133;  Gaps   30;

Qy        206 PADPDWIAFIHSND----IWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEE- 260
              |  :  ||  :         :| ::|  :  |  |    |   :  :|:    |:  :| :||

Db        147 PNNTQWITWSQEGHKLAYVWKNDIYVKIEPHLP-SHRTTSTGKENVIFNGINDWVYEEEI 205

Qy        261 FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGT 320
              |  ||  ||       :|:|  :    : :      :|    :   |:  :      ||| |
Db        206 FGAYSALWW------SPNGTFLAYAQFNDTGVPLIEYSFYSDESLQYPKTVWIPYPKAGA 259

Qy        321 ANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLD 380
              || |*| :  :  |:       |  ::  |  :   |  |:    |  | : :
Db        260 VNPTVKFFI--VNTDSLSSTTTTIPMQITAPASVT-TGDHYLCDVAWVSEDRIS------ 310

Qy        381 RSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWININHDIFHV- 439
                                    ::  | |::   ||  :  |::|  :|       |:
Db        311 ----------------------LQWLRRIQNY--SVMAICDYDKTNLVWNCPTTQEHIE 345

Qy        440 -------------FPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPS 486
                          |    :   |       | |::|   |     |: |
Db        346 TSATGWVGRFRPAEPHFTSDGSSFYKIVSDKDGYKH---ICQFQKDRK------------ 390

Qy        487 DFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGT-KDSPLEHHLYVVSYVNPG 545
                | ::   || | |||:     :|:       | |      |: |   :|| :
Db        391 ----PEQDCTFITKGAWEVI-----SIEALTSDYLYYISNEYKEMPGGRNLYKI------ 435

Qy        546 EVTRLTDRGYSHSCCIS-----QHCDFF---ISKYSNQKNPHC----VSLYKLSSPEDDP 593
                :|||  :::  |:|      : | ::   :|| :       |     : || |     |
Db        436 ---QLTD--HTNKKCLSCDLNPERCQYYSVSLSKEAKYYQLGCRGPGLPLYTLHRSTDQK 490

Qy        594 TCKTKEFWATILDSAGPLPDYTPPE---IFSFESTTGFTLYGMLYKPHDLQPGKKYPTVL 650
              :  |   : ||     | |   |      |   : | |  | |:  ||      |||| ::
Db        491 ELRVLED-NSALDKM--LQDVQMPSKKLDFIVLNETRF-WYQMILPPH-FDKSKKYPLLI 545

Qy        651 FIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLKFEGAFKYKM 706
              :| ||  |      |   || |   |||   ::|   | ||| ::| |   |   ::
Db        546 DVYAGPCSQ------KADAAFRFNWATYLASTENIIVASFDGRGSGYQGDKIMHAINKRL 599

Qy        707 GQIEIDDQVEGL-QYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPV 765
              | :|::||:|   |:|  :  |:|  :| | |||||||:: | |    | :|:  || |||
Db        600 GTLEVEDQIEAARQFL--KMGFVDSKQVAIWGWSYGGYVTSMVLGSGSGVFKCGIAVAPV 657

Qy        766 TLWIFYDTGYTERYMG--HPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFA 823
              : | :||: |||||||    |: |   |   :|  :|| |  :     ||:||  |:||||
Db        658 SRWEYYDSVYTERYMGLPTPEDNLDHYRNSTVMSRAENF--KQVEYLLIHGTADDNVHFQ 715

Qy        824 HTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQE 870
              :: :   || ||  :   |  | | |    : :|   |: |:||:
Db        716 QSAQISKALVDAGVDFQAMWYTDEDHGIASSTAHQHIYSHMSHFLQQ 762
```

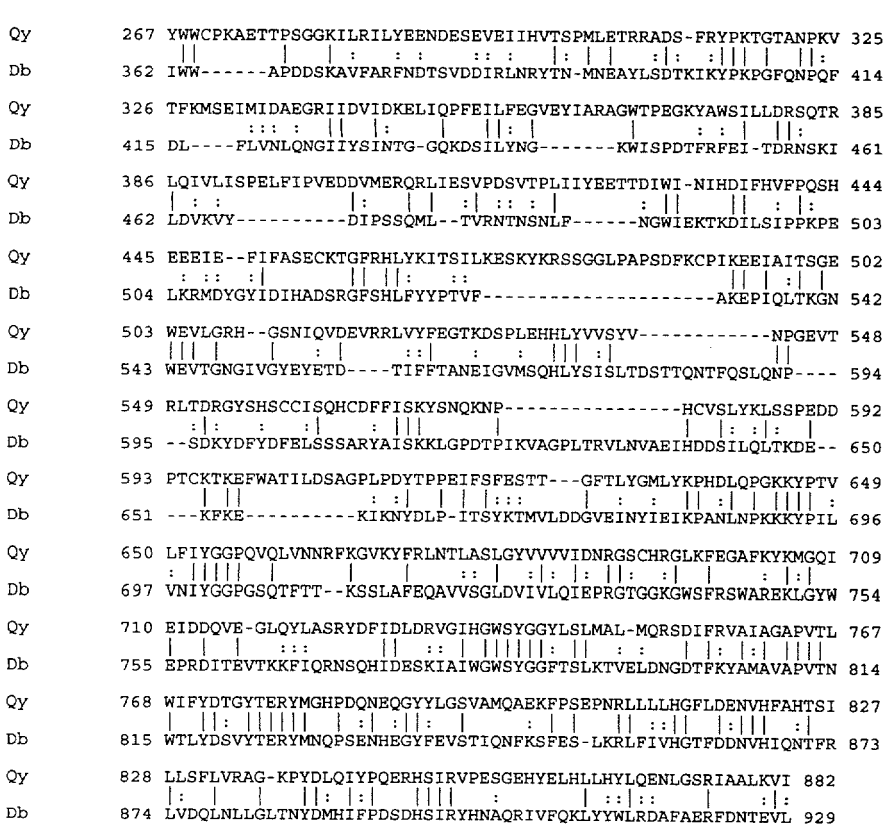
```
RESULT 9
A30107
dipeptidyl aminopeptidase B (EC 3.4.14.-) - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YHR028c
C;Species: Saccharomyces cerevisiae
C;Date: 07-Jun-1990 #sequence_revision 30-May-1997 #text_change 23-Mar-2001
C;Accession: S46780; A30107
R;Du, Z.
submitted to the EMBL Data Library, June 1994
A;Description: The sequence of S. cerevisiae cosmid 8082.
A;Reference number: S46773
A;Accession: S46780
A;Molecule type: DNA
A;Residues: 1-818 <DUZ>
A;Cross-references: EMBL:U10399; NID:g500689; PIDN:AAB68879.1; PID:g500698; GSPDB:GN000
R;Roberts, C.J.; Pohlig, G.; Rothman, J.H.; Stevens, T.H.
J. Cell Biol. 108, 1363-1373, 1989
A;Title: Structure, biosynthesis, and localization of dipeptidyl aminopeptidase B, an i
A;Reference number: A30107; MUID:89174971; PMID:2647766
A;Accession: A30107
A;Molecule type: DNA
A;Residues: 1-82,'H',84-124,'N',126-181,'LRRLET',189-199,'N',201-365,'DFKRGKERKF',376-5
A;Cross-references: EMBL:X15484
A;Note: the authors translated the codon ACC for residue 572 as Asn
C;Genetics:
A;Gene: SGD:DAP2; STE13; MIPS:YHR028c
```

A;Cross-references: SGD:S0001070; MIPS:YHR028c
A;Map position: 8R
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein; yeast vacu
F;30-45/Domain: transmembrane #status predicted <TMM>
F;63,79,110,139,392,421/Binding site: carbohydrate (Asn) (covalent) #status predicted

```
  Query Match             10.3%;  Score 484.5;  DB 1;  Length 818;
  Best Local Similarity   26.0%;  Pred. No. 2.4e-26;
  Matches  183;  Conservative   95;  Mismatches  250;  Indels  177;  Gaps   35;

Qy        212 IAFIHSNDIWISNIVTREERRLTYVHNE----LANMEEDARSAGVATFVLQEE-FDRYSG 266
              ||::  |:|:| : ::::  |    | |:     | | :  |        :| :|| |:
Db        201 IAYVQDNNIYIYSAISKKTIRA--VTNDGSSFLFNGKPD--------WVYEEEVFEDDKA 250

Qy        267 YWWCPKAETTPSGGKILRILYEENDESEV-EIIHVTSPMLETRRAD------SFRYPKTG 319
               ||  |   :            : : : ||||| | |     |         |      | :|||:|
Db        251 AWWSPTGD---------YLAFLKIDESEVGEFI---IPYYVQDEKDIYPEMRSIKYPKSG 298

Qy        320 TANPKVTFKMSEIMIDAEGRIIDVIDKELIQP--FEILFEGVEYIARAGWTPEGKYAWSI 377
              | ||                 ||  :  : |     |          :|   |     |     |       ::
Db        299 TPNP-----------HAELWVYSMKDGTSFHPRISGNKKDGSLLITEVTWVGNG----NV 343

Qy        378 LL---DRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINI- 433
              |:    |||   |: ||            |  :      : :              |        | :
Db        344 LVKTTDRSSDILTVFLI---------DTIAKTSNVVRN-----------ESSNGGWWEIT 383

Qy        434 HDIFHV-----FPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDF 488
              |:   :       |  : |     :: :            |: ||           | :|         | :
Db        384 HNTLFIPANETFDRPHNGYVDILPIG----GYNHL----------AYFENSNS----SHY 425

Qy        489 KCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVT 548
              |           :| |:|||:        :    |  :     :||    |:  |    | |:|  :    :| |:
Db        426 K-------TLTEGKWEVV---NGPLAFDSMENRLYFISTRKSSTERHVYYIDLRSPNEII 475

Qy        549 RLTDRGYS--HSCCISQHCDFFISKYSNQKNPH-------------C-------VSLYKL 586
               :||        :      |       | :  |    | |:                    |         ||| |
Db        476 EVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHL 535

Qy        587 SSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYG--MLYKPHDLQP-- 642
                      |              :|   ||  |         |        |     :|    :|     ::: |
Db        536 EKNE-------------VLTKI--LEDYAVPRKSFRELNLGKDEFGKDILVNSYEILPND 580

Qy        643 -----GKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN----TLASLGYVVVVIDNRGSCH 693
                        ||    |  ||||  | |            | |  :          :   |  :|||:|  ||:
Db        581 FDETLSDHYPVFFFAYGGPNSQQV------VKTFSVGFNEVVASQLNAIVVVVDGRGTGF 634

Qy        694 RGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRS 753
               :|   |       : ::|    |    ||:           |      |:|   :: :  |||||||:|     |   :
Db        635 KGQDFRSLVRDRLGDYEARDQISAASLYGS-LTFVDPQKISLFGWSYGGYLTLKTLEKDG 693

Qy        754 D-IFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSV----AMQAEKFPSEPNR 808
                 |:    ::  |||| | |||: ||||||   |  :|    ||     ||       |:              :: ||
Db        694 GRHFKYGMSVAPVTDWRFYDSVYTERYMHTPQENFDGYVESSVHNVTAL------AQANR 747

Qy        809 LLLLHGFLDENVHFAHTSILLSFLVRAG-KPYDLQIYPQERHSIR 852
               ||:||   |:||||  ::     |    |        |  :  ||: ::|        ||||
Db        748 FLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIR 792
```

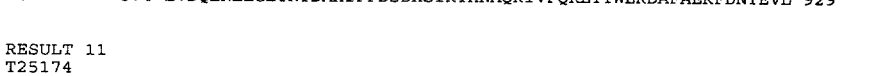
RESULT 10
A49737
dipeptidyl aminopeptidase (EC 3.4.14.-) - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein O5045; protein YOR219c; protein YOR50-9
C;Species: Saccharomyces cerevisiae
C;Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 29-Oct-1999
C;Accession: A49737; S45451; S60946; S67112; S71721
R;Santa Anna-A, S.; Herskowitz, I.
submitted to the Protein Sequence Database, July 1993
A;Reference number: A49737
A;Accession: A49737
A;Molecule type: DNA
A;Residues: 1-931 <SAN>
A;Cross-references: GB:L21944; NID:g347196; PIDN:AAA35119.1; PID:g347197
R;Anna-Arriola, S.S.; Herskowitz, I.
Yeast 10, 801-810, 1994
A;Title: Isolation and DNA sequence of the STE13 gene encoding dipeptidyl aminopeptidas
A;Reference number: S45451; MUID:95066382; PMID:7975897
A;Accession: S45451
A;Molecule type: DNA
A;Residues: 1-931 <ANN>
A;Cross-references: EMBL:L21944; NID:g347196; PIDN:AAA35119.1; PID:g347197
R;Galisson, F.; Dujon, B.
submitted to the EMBL Data Library, October 1995
A;Description: Sequence and analysis of a 33 kb fragment from the right arm of chromoso
A;Reference number: S60938
A;Accession: S60946
A;Molecule type: DNA
A;Residues: 1-931 <GAL>
A;Cross-references: EMBL:X92441; NID:g1050762; PIDN:CAA63182.1; PID:g1050771
R;Boyer, J.; Fairhead, C.; Gaillon, L.; Galisson, F.; Michaux, G.; Thierry, A.; Dujon,
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67104
A;Accession: S67112
A;Molecule type: DNA
A;Residues: 1-931 <BOY>
A;Cross-references: EMBL:Z75127; NID:g1420507; PIDN:CAA99437.1; PID:e252398; PID:g14205
A;Experimental source: strain S288C
R;Galisson, F.; Dujon, B.
Yeast 12, 877-885, 1996
A;Title: Sequence and analysis of a 33 kb fragment from the right arm of chromosome XV
A;Reference number: S71713; MUID:96437977; PMID:8840505
A;Accession: S71721
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-931 <GAW>
A;Cross-references: EMBL:X92441; NID:g1050762; PIDN:CAA63182.1; PID:g1050771
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995
C;Genetics:
A;Gene: SGD:STE13; YCI1
A;Cross-references: SGD:S0005745; MIPS:YOR219c
A;Map position: 15R
C;Function:
A;Description: involved in processing of alpha-factor prepropheromone
C;Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein
F;113-150/Domain: transmembrane #status predicted <TMM>
F;377/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;785,863,896/Active site: Ser, Asp, His #status predicted

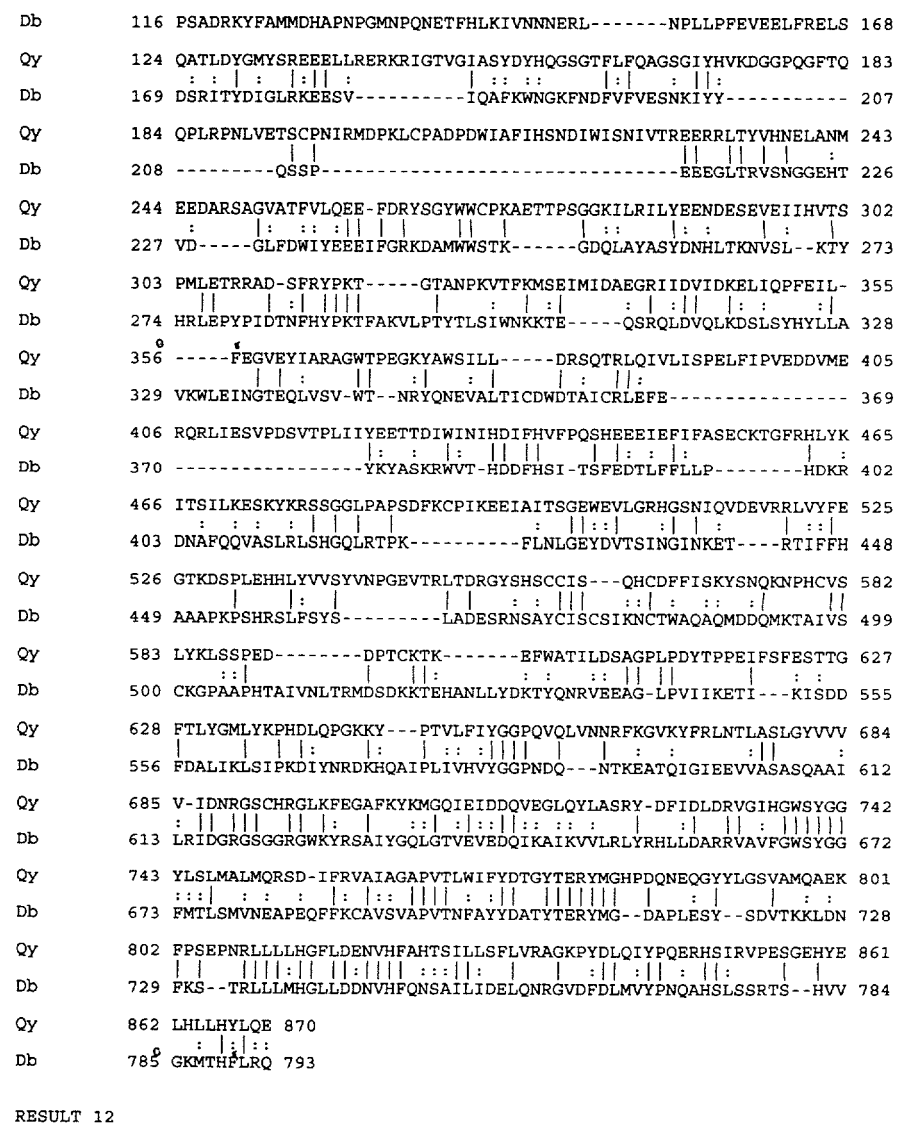
```
  Query Match              9.9%;  Score 465;  DB 2;  Length 931;
  Best Local Similarity   21.5%;  Pred. No. 7.4e-25;
  Matches  218;  Conservative  150;  Mismatches  380;  Indels  268;  Gaps   41;

Qy          6 ETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMMAKAPH 65
               :: :  |  |    |   |:   |:  ||  |:||                                 |
Db         43 QSNENGHTINEIRPTEATIDVTDVPQ-TPFLQEQYS-----------------MRPRRE 83

Qy         66 DFMFVKRNDPDGPHSDRIYYLAMS----------GENRENTLFYSEIPKTINRAAVLMLS 115
               |  |    || :   |      ::                   |  |   | :: |   ::     |:::
Db         84 SFQF---NDIENQHHTHSFFSVNKFNRRWGEWSLPEKRSYVLVFTLIALSVLVLLVILIP 140

Qy        116 WKPL----------------------------------------LDLFQATL----DYGM 131
               |  | |                                       :|  |   |      | |:
Db        141 SKLLPTKITRPKTSAGDSSLGKRSFSIENVLNGDFAIPEDTFHFIDPPQRLLGQDSDPGL 200

Qy        132 YSREEELLRERKRIG-TVGIASYDYHQGSGTFLFQ------------------------ 165
               |    :|:       |    :    ::: :  |   ||::
Db        201 YFTTKEIDGHTNFIAKQLFDETFEVNLGGNRFLYEGVEFTVSTVQINYKLDKLIFGTNLE 260

Qy        166 -----AGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPK------------LCPAD 208
                    :  |  | :||           |     :|  |   :  |              ||
Db        261 SEFRHSSKGFYWIKD----------LNTGNIEPILPPEKSDDNYELGLSKLSYAHFSPA- 309

Qy        209 PDWIAFIHSNDIWISNIVTREERRLTY-VHNELANMEEDARSAGVATFVLQEE-FDRYSG 266
               ::|  |:: |::::  :  :    :::|                         :: :||
Db        310 YNYIYFVYENNLFLQQVNSGVAKKVTEDGSKDIFNAKPD--------WIYEEEVLASDQA 361
```



RESULT 11
T25174
hypothetical protein T23F1.7b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C;Accession: T25174
R;Wilkinson, J.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19990
A;Accession: T25174
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-799 <WIL>
A;Cross-references: EMBL:Z81129; PIDN:CAB03412.1; GSPDB:GN00023; CESP:T23F1.7b
A;Experimental source: clone T23F1
C;Genetics:
A;Gene: CESP:T23F1.7b
A;Map position: 5
A;Introns: 13/3; 52/1; 111/3; 197/3; 366/3; 524/1; 557/1; 648/2; 703/1; 743/1
C;Superfamily: dipeptidyl-peptidase IV

Query Match 9.6%; Score 451.5; DB 2; Length 799;
Best Local Similarity 23.0%; Pred. No. 5.4e-24;
Matches 195; Conservative 129; Mismatches 298; Indels 227; Gaps 38;





RESULT 12
T25173
hypothetical protein T23F1.7a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C;Accession: T25173
R;Wilkinson, J.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19990
A;Accession: T25173
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-779 <WIL>
A;Cross-references: EMBL:Z81129; PIDN:CAB03411.1; GSPDB:GN00023; CESP:T23F1.7a

```
A;Experimental source: clone T23F1
C;Genetics:
A;Gene: CESP:T23F1.7a
A;Map position: 5
A;Introns: 13/3; 52/1; 111/3; 151/2; 177/3; 346/3; 504/1; 537/1; 628/2; 683/1; 723/1
C;Superfamily: dipeptidyl-peptidase IV

  Query Match             9.5%;  Score 445.5;  DB 2;  Length 779;
  Best Local Similarity   23.0%;  Pred. No. 1.4e-23;
  Matches  193;  Conservative  126;  Mismatches  292;  Indels  229;  Gaps   38;

Qy         78 PHSDRIYYLAMS-----GENRENTLFYSEIPKTINRAAVLMLSWKPLLDLFQATLDYGMY 132
Db        116 PSADRKYFAMMDHAPNPGMNPQNETFHLKIVNNNER----------------ITYDIGL- 158

Qy        133 SREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVE 192
Db        159 -RKEESV----------IQAFKWNGKFNDFVFVESNKIYY-------------------- 187

Qy        193 TSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGV 252
Db        188 QSSP-------------------------------EEEGLTRVSNGGEHTVD-----GL 210

Qy        253 ATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLETRRAD 311
Db        211 FDWIYEEEIFGRKDAMWWSTK------GDQLAYASYDNHLTKNVSL--KTYHRLEPYPID 262

Qy        312 -SFRYPKT-----GTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEIL------FEGV 359
Db        263 TNFHYPKTFAKVLPTYTLSIWNKKTE-----QSRQLDVQLKDSLSYHYLLAVKWLEINGT 317

Qy        360 EYIARAGWTPEGKYAWSILL-----DRSQTRLQIVLISPELFIPVEDDVMERQRLIESVP 414
Db        318 EQLVSV-WT--NRYQNEVALTICDWDTAICRLEFE------------------------- 349

Qy        415 DSVTPLIIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESK 474
Db        350 --------YKYASKRWVT-HDDFHSI-TSFEDTLFFLLP--------HDKRDNAFQQVAS 391

Qy        475 YKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEH 534
Db        392 LRLSHGQLRTPK----------FLNLGEYDVTSINGINKET----RTIFFHAAAPKPSHR 437

Qy        535 HLYVVSYVNPGEVTRLTDRGYSHSCCIS---QHCDFFISKYSNQKNPHCVSLYKLSSPED 591
Db        438 SLFSYS---------LADESRNSAYCISCSIKNCTWAQAQMDDQMKTAIVSCKGPAAPHT 488

Qy        592 --------DPTCKTK-------EFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYK 636
Db        489 AIVNLTRMDSDKKTEHANLLYDKTYQNRVEEAG-LPVIIKETI---KISDDFDALIKLSI 544

Qy        637 PHDLQPGKKY---PTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVV-IDNRGSC 692
Db        545 PKDIYNRDKHQAIPLIVHVYGGPNDQ---NTKEATQIGIEEVVASASQAAILRIDGRGSG 601

Qy        693 HRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRY-DFIDLDRVGIHGWSYGGYLSLMALMQ 751
Db        602 GRGWKYRSAIYGQLGTVEVEDQIKAIKVVLRLYRHLLDARRVAVFGWSYGGFMTLSMVNE 661

Qy        752 RSD-IFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLL 810
Db        662 APEQFFKCAVSVAPVTNFAYYDATYTERYMG--DAPLESY--SDVTKKLDNFKS--TRLL 715

Qy        811 LLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQE 870
Db        716 LMHGLLDDNVHFQNSAILIDELQNRGVDFDLMVYPNQAHSLSSRTS--HVVGKMTHFLRQ 773
```

RESULT 13
A41793
dipeptidyl aminopeptidase-like protein (EC 3.4.14.-) short form, DPPX-S - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999

```
C;Accession: A41793
R;Wada, K.; Yokotani, N.; Hunter, C.; Doi, K.; Wenthold, R.J.; Shimasaki, S.
Proc. Natl. Acad. Sci. U.S.A. 89, 197-201, 1992
A;Title: Differential expression of two distinct forms of mRNA encoding members of a di
A;Reference number: A41793; MUID:92108018; PMID:1729689
A;Accession: A41793
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-803 <WAD>
A;Cross-references: GB:M76429; NID:g408719; PIDN:AAC41623.1; PID:g408720
A;Note: sequence extracted from NCBI backbone (NCBIP:75138)
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptide hydrolase; glycoprotein
F;257,342/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             9.4%;  Score 440;  DB 2;  Length 803;
  Best Local Similarity   24.4%;  Pred. No. 3.6e-23;
  Matches  176;  Conservative  114;  Mismatches  317;  Indels  114;  Gaps   29;

Qy        171 YHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREE 230
Db        156 YYVLSKIPHGDPQSLDPPEVSNAKLQYAGWGPK-----GQQLIFIFENNIYYCAHVGKQA 210

Qy        231 RRLTYVHNELANMEEDARSAGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEE 289
Db        211 IRV------VSTGKEGVIYNGLSDWLYEEEILKTHIAHWWSPDG---------TRLAYAT 255

Qy        290 NDESEVEIIHV---TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDK 346
Db        256 INDSRVPVMELPTYTGSVYPT--AKPYHYPKAGCENPSISLH----VIGLNGPTHDL--- 306

Qy        347 ELIQPFEILFEGVEYIARAGWTPEGKYA--WSILLDRSQTRLQIVLISPELFIPVEDDVM 404
Db        307 EMTPPDDPRMREY-YITMVKWATSTKVAVNW---LSRAQN------VS----------- 344

Qy        405 ERQRLIESVPDSVTPLII--YEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRH 462
Db        345 -----ILTLCDATTGVCTKKHEDESEAWL------------HRQNEEPVFS---KDG-RK 383

Qy        463 LYKITSILK--ESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRR 520
Db        384 FFFVRAIPQGGQGKFYHITVSSSQPNSSNDNIQ---SITSGDWDVT----KILSYDEKRS 436

Qy        521 LVYFEGTKDSPLEHHLYVVSYVNPGEVTR------LTDRGYSHSCCISQHCDFFISKYSN 574
Db        437 QIYFLSTEDLPRRRQLYSASTV--GSFNRQCLSCDLVDNCTYFSASFSPGADFFLLKCEG 494

Qy        575 QKNPHCVSLYKLSSPEDDPTCKTKE-FWATILDSAGPLPDYTPPEIFSFESTTGFTLYGM 633
Db        495 PGVP-TVSVHNTTDKKKMFDLETNEHVQKAISDRQMPKVEYRKIE------TDDYNLPIQ 547

Qy        634 LYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCH 693
Db        548 ILKPATFTDTAHYPLLLVVDGTPGSQSVAEKF--AVTWETVMVSSHGAVVVKCDGRGSGF 605

Qy        694 RGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRS 753
Db        606 QGTRLLHEVRRRLGSLEEKDQMEAVRVML-KEPYIDKTRVAVFGKDYGGYLSTYLLPAKG 664

Qy        754 D----IFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRL 809
Db        665 DGQAPVFSCGSALSPITDFKLYASAFSERYLGLHGLDNRAYEMAKVAHRVSAL--EGQQF 722

Qy        810 LLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQ 869
Db        723 LVIHATADEKIHFQHTAELITQLIKGKANYSLQIYPDESHYFSSAALQQHLHRSILGFFV 782

Qy        870 E 870
Db        783 E 783
```

RESULT 14

I68600
dipeptidyl aminopeptidase like protein - human
C;Species: Homo sapiens (man)
C;Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 21-Jul-2000
C;Accession: I68600
R;Yokotani, N.; Doi, K.; Wenthold, R.J.; Wada, K.
Hum. Mol. Genet. 2, 1037-1039, 1993
A;Title: Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-relat
A;Reference number: I54331; MUID:93372805; PMID:8103397
A;Accession: I68600
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-803 <RES>
A;Cross-references: GB:M96860; NID:g306707; PIDN:AAA35761.1; PID:g306708
C;Superfamily: dipeptidyl-peptidase IV

Query Match 9.1%; Score 427; DB 2; Length 803;
Best Local Similarity 23.4%; Pred. No. 3.1e-22;
Matches 172; Conservative 121; Mismatches 299; Indels 144; Gaps 30;

```
Qy      171 YHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREE 230
Db      156 YYVLSKIPHGDPQSLDPPEVSNAKLQYAGWGPK-----GQQLIFIFENNIYYCAHVGKQA 210

Qy      231 RRLTYVHNELANMEEDARSAGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEE 289
Db      211 IRV------VSTGKEGVIYNGLSDWLYEEEILKTHIAHWWSPDG---------TRLAYAA 255

Qy      290 NDESEVEIIHV---TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDK 346
Db      256 INDSRVPIMELPTYTGSIYPT--VKPYHYPKAGSENPSISLH----VIGLNGPTHDL--- 306

Qy      347 ELIQPFEILFEGVEYIARAGWTPEGKYA--WSILLDRSQTRLQIVLISPELFIPVEDDVM 404
Db      307 EMMPPDDPRMREY-YITMVKWATSTKVAVTW---LNRAQN------VS------------ 344

Qy      405 ERQRLIESVPDSVTPLII--YEETTDIWINIHDIFHVFPQSHEEEI------EFIFASEC 456
Db      345 -----ILTLCDATTGVCTKKHEDESEAWLH---------RQNEEPVFSKDGRKFFFIRAI 390

Qy      457 KTGFR-HLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQV 515
Db      391 PQGGRGKFYHIT--VSSSQPNSSNDNIQ------------SITSGDWDVT----KILAY 431

Qy      516 DEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSH---SCCISQHCDFFISKY 572
Db      432 DEKGNKIYFLSTEDLPRRRQLYSAN----------TEGNFNRQCLSCDLVENCTYFSASF 481

Qy      573 SNQKNPHCVSLYKLSSP------EDDPTCKTKEF--------WATILDSAGPLPDYTPPE 618
Db      482 SHSMD---FFLLKCEGPGVPMVTVHNTTDKKKMFDLETNEHVKKAINDRQMPKVEYRDIE 538

Qy      619 IFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLAS 678
Db      539 I------DDYNLPMQILKPATFTDTTHYPLLLVVDGTPGSQSVAEKFE--VSWETVMVSS 590

Qy      679 LGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGW 738
Db      591 HGAVVVKCDGRGSGFQGTKLLHEVRRRLGLLEEKDQMEAVRTML-KEQYIDRTRVAVFGK 649

Qy      739 SYGGYLSLMALMQRSD----IFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGS 794
Db      650 DYGGYLSTYILPAKGENQGQTFTCGSALSPITDFKLYASAFSERYLGLHGLDNRAYEMTK 709

Qy      795 VAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVP 854
Db      710 VAHRVSAL--EEQQFLIIHPTADEKIHFQHTAELITQLIRGKANYSLQIYPDESHYFTSS 767

Qy      855 ESGEHYELHLLHYLQE 870
Db      768 SLKQHLYRSIINFFVE 783
```

RESULT 15
I54331
dipeptidyl aminopeptidase like protein - human
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C;Accession: I54331
R;Yokotani, N.; Doi, K.; Wenthold, R.J.; Wada, K.
Hum. Mol. Genet. 2, 1037-1039, 1993
A;Title: Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-rela
A;Reference number: I54331; MUID:93372805; PMID:8103397
A;Accession: I54331
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-865 <RES>
A;Cross-references: GB:M96859; NID:g306705; PIDN:AAA35760.1; PID:g306706
C;Superfamily: dipeptidyl-peptidase IV

Query Match 9.1%; Score 427; DB 2; Length 865;
Best Local Similarity 23.4%; Pred. No. 3.4e-22;
Matches 172; Conservative 121; Mismatches 299; Indels 144; Gaps 30;

```
Qy      171 YHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREE 230
Db      218 YYVLSKIPHGDPQSLDPPEVSNAKLQYAGWGPK-----GQQLIFIFENNIYYCAHVGKQA 272

Qy      231 RRLTYVHNELANMEEDARSAGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEE 289
Db      273 IRV------VSTGKEGVIYNGLSDWLYEEEILKTHIAHWWSPDG---------TRLAYAA 317

Qy      290 NDESEVEIIHV---TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDK 346
Db      318 INDSRVPIMELPTYTGSIYPT--VKPYHYPKAGSENPSISLH----VIGLNGPTHDL--- 368

Qy      347 ELIQPFEILFEGVEYIARAGWTPEGKYA--WSILLDRSQTRLQIVLISPELFIPVEDDVM 404
Db      369 EMMPPDDPRMREY-YITMVKWATSTKVAVTW---LNRAQN------VS------------ 406

Qy      405 ERQRLIESVPDSVTPLII--YEETTDIWINIHDIFHVFPQSHEEEI------EFIFASEC 456
Db      407 -----ILTLCDATTGVCTKKHEDESEAWLH---------RQNEEPVFSKDGRKFFFIRAI 452

Qy      457 KTGFR-HLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQV 515
Db      453 PQGGRGKFYHIT--VSSSQPNSSNDNIQ------------SITSGDWDVT----KILAY 493

Qy      516 DEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSH---SCCISQHCDFFISKY 572
Db      494 DEKGNKIYFLSTEDLPRRRQLYSAN----------TEGNFNRQCLSCDLVENCTYFSASF 543

Qy      573 SNQKNPHCVSLYKLSSP------EDDPTCKTKEF--------WATILDSAGPLPDYTPPE 618
Db      544 SHSMD---FFLLKCEGPGVPMVTVHNTTDKKKMFDLETNEHVKKAINDRQMPKVEYRDIE 600

Qy      619 IFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLAS 678
Db      601 I------DDYNLPMQILKPATFTDTTHYPLLLVVDGTPGSQSVAEKFE--VSWETVMVSS 652

Qy      679 LGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGW 738
Db      653 HGAVVVKCDGRGSGFQGTKLLHEVRRRLGLLEEKDQMEAVRTML-KEQYIDRTRVAVFGK 711

Qy      739 SYGGYLSLMALMQRSD----IFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGS 794
Db      712 DYGGYLSTYILPAKGENQGQTFTCGSALSPITDFKLYASAFSERYLGLHGLDNRAYEMTK 771

Qy      795 VAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVP 854
Db      772 VAHRVSAL--EEQQFLIIHPTADEKIHFQHTAELITQLIRGKANYSLQIYPDESHYFTSS 829

Qy      855 ESGEHYELHLLHYLQE 870
Db      830 SLKQHLYRSIINFFVE 845
```

Search completed: October 15, 2003, 17:14:45
Job time : 31.176 secs

OM protein - protein search, using sw model

Run on: October 15, 2003, 16:57:30 ; Search time 13.9931 Seconds
(without alignments)
2964.152 Million cell updates/sec

Title: US-10-070-464-1
Perfect score: 4700
Sequence: 1 MAAAMETEQLGVEIFETADC.....HLLHYLQENLGSRIAALKVI 882

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	569.5	12.1	765	1	DPP4_FELCA	Q9n2i7 felis silve
2	566	12.0	765	1	DPP4_BOVIN	P81425 bos taurus
3	525	11.2	766	1	DPP4_HUMAN	P27487 homo sapien
4	523.5	11.1	767	1	DPP4_RAT	P14740 rattus norv
5	519.5	11.1	793	1	YEA8_SCHPO	O14073 schizosacch
6	517.5	11.0	760	1	DPP4_MOUSE	P28843 mus musculu
7	484.5	10.3	818	1	DAP2_YEAST	P18962 saccharomyc
8	478.5	10.2	761	1	SEPR_MOUSE	P97321 mus musculu
9	465	9.9	931	1	ST13_YEAST	P33894 saccharomyc
10	454	9.7	760	1	SEPR_HUMAN	Q12884 homo sapien
11	440	9.4	863	1	DPP6_BOVIN	P42659 bos taurus
12	435	9.3	804	1	DPP6_MOUSE	Q9z218 mus musculu
13	431	9.2	859	1	DPP6_RAT	P46101 rattus norv
14	427	9.1	865	1	DPP6_HUMAN	P42658 homo sapien
15	293	6.2	853	1	YDZF_SCHPO	Q9p7e9 schizosacch
16	249.5	5.3	657	1	YUXL_BACSU	P39839 bacillus su
17	172	3.7	732	1	ACPH_HUMAN	P13798 homo sapien
18	171	3.6	732	1	ACPH_RAT	P13676 rattus norv
19	161	3.4	732	1	ACPH_PIG	P19205 sus scrofa
20	150	3.2	740	1	YL31_CAEEL	P34422 caenorhabdi
21	139	3.0	689	1	PPCE_AERHY	Q06903 aeromonas h
22	134	2.9	722	1	Y174_RICPR	P81171 rickettsia
23	134	2.9	1014	1	BGAL_BACHD	Q9k9c6 bacillus ha
24	125.5	2.7	436	1	PAFA_CAVPO	P70683 c platelet-
25	124.5	2.6	1835	1	DUR1_YEAST	P32528 saccharomyc
26	123.5	2.6	705	1	PPCF_FLAME	P27195 flavobacter
27	123	2.6	1702	1	DPOL_THELI	P30317 thermococcu
28	122	2.6	705	1	Y4SO_RHISN	P55656 rhizobium s
29	119.5	2.5	1019	1	CA26_HUMAN	P12110 homo sapien
30	118.5	2.5	705	1	PPCE_FLAME	P27028 flavobacter
31	118.5	2.5	710	1	PPCE_HUMAN	P48147 homo sapien
32	118.5	2.5	1436	1	DPO3_STAAW	Q53665 staphylococ
33	118.5	2.5	1438	1	DPO3_STAAM	Q99uk8 staphylococ
34	118	2.5	1285	1	TOXA_PASMU	P17452 pasteurella
35	115	2.4	1047	1	POL_SIVA1	P27973 simian immu
36	115	2.4	3144	1	VP13_YEAST	Q07878 saccharomyc
37	114.5	2.4	710	1	PPCE_MOUSE	Q9qur6 mus musculu
38	113.5	2.4	2483	1	COA2_HUMAN	O00763 homo sapien
39	112.5	2.4	835	1	NLG2_HUMAN	Q8nfz4 homo sapien
40	111.5	2.4	465	1	EGLC_RHIME	Q9z3q2 rhizobium m
41	110.5	2.4	836	1	NLG2_RAT	Q62888 rattus norv
42	110.5	2.4	4349	1	FAT2_HUMAN	Q9nyq8 homo sapien
43	110	2.3	1295	1	BXA1_CLOBO	P10845 clostridium
44	110	2.3	1861	1	MAP2_RAT	P15146 rattus norv
45	110	2.3	2410	1	MOK1_SCHPO	Q9usk8 schizosacch

ALIGNMENTS

RESULT 1
DPP4_FELCA
ID DPP4_FELCA STANDARD; PRT; 765 AA.
AC Q9N2I7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26).
GN DPP4 OR CD26.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood;
RX MEDLINE=20094000; PubMed=10630304;
RA Nishimura Y., Miyazawa T., Ikeda Y., Izumiya Y., Nakamura K., Sato E.,
RA Mikami T., Takahashi E.;
RT "Molecular cloning and sequencing of a cDNA encoding the feline T-cell
RT activation antigen CD26 homologue.";
RL Immunogenetics 50:366-368(1999).
CC -!- FUNCTION: Removes N-terminal dipeptides sequentially from
CC polypeptides having unsubstituted N-termini provided that the
CC penultimate residue is proline (By similarity).
CC -!- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC Xcc is neither Pro nor hydroxyproline.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC a soluble form (By similarity).
CC -!- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC by proteolytic processing (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family S9B.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AB023952; BAA92344.1; -.
DR MEROPS; S09.003; -.
DR InterPro; IPR002469; DPPIV_N_term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002471; Prol_endopep_ser.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00930; DPPIV_N_term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW Transmembrane; Glycoprotein; Signal-anchor.
FT CHAIN 1 765 DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM

```
FT                                     (MDPP).
FT   CHAIN        38    765            DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT                                     (SDPP) (BY SIMILARITY).
FT   DOMAIN        1      6            CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     29            SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                     (POTENTIAL).
FT   DOMAIN       30    765            EXTRACELLULAR (POTENTIAL).
FT   ACT_SITE    629    629            CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    707    707            CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    739    739            CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     84     84            N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     91     91            N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    149    149            N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    178    178            N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    228    228            N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    280    280            N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    320    320            N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    330    330            N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    331    331            N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    519    519            N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    684    684            N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   765 AA;  88213 MW;  3EFCE98A22B175D9 CRC64;

  Query Match             12.1%;  Score 569.5;  DB 1;  Length 765;
  Best Local Similarity   25.7%;  Pred. No. 8.2e-34;
  Matches  225;  Conservative  114;  Mismatches  316;  Indels  221;  Gaps   42;

Qy         71 KRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLLDLFQATLDYG 130
              | ||      | | | |    :|     |||           | |       :  ||
Db         30 KGNDA-AADSRRTYTLTDYLKNTFRVKFYS-------------LRW-------VSDHDY- 67

Qy        131 MYSREEELLRERKRIGTVGI----ASYDYHQGS---------GTFL-------------F 164
              :| ::  :|       |    |    :::|  : |         | |:            :
Db         68 LYKQDNNILLFNAEYGNSSIFLENSTFDEFEHSINDYSVSPDGQFILLEYNYVKQWRHSY 127

Qy        165 QAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWI---------AFI 215
              |   || :               |  : |   ||          :  ||         |::
Db        128 TASYDIYDLNK-----------RQLITEEKIPN----------NTQWITWSPEGHKLAYV 166

Qy        216 HSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEE-FDRYSGYWWCPKAE 274
              ||::: |       |:|:        ||:|   |:| :| :|| |  ||  || ||
Db        167 WKNDVYVKNEPNSSSHRITWTG------EENAIYNGIADWVYEEEIFSAYSALWWSPKGT 220

Qy        275 TTPSGGKILRILYEENDESEVEIIH---VTSPMLETRRADSFRYPKTGTANPKVTFKMSE 331
                : | : ::::| :|     :   |:          ||| | |||  | |:
Db        221 F---------LAYAQFNDTQVPLIEYSFYSDESLQYPMTMRIPYPKAGAANP--TVKLFV 269

Qy        332 IMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLI 391
              |  |      :    |:  |  :|  |  |:    |  | : :
Db        270 IKTDNLNPNTNATSVEITPPAAML-TGDYYLCDVTWANEERIS---------------- 311

Qy        392 SPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSHEEEIEFI 451
                        ::  | |::   ||  :  |  :|  ||:          : :| ||
Db        312 -----------LQWLRRIQNY--SVMDIRDYNNSTGKWIS---------SAAQEHIEM- 347

Qy        452 FASECKTGF---------------RHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEI 496
                ||:                 |: ||| |   |  ||          |:   |:
Db        348 ----STTGWVGRFRPAEPHFTSDGRNFYKIIS--NEDGYKHI-------CRFQIDKKDCT 394

Qy        497 AITSGEWEVLGRHGSNIQVDEVRRLVYFEGT-KDSPLEHHLYVVSYVNPGEVTRLTDRGY 555
              || | |||:|     |:      | |     |  |   :|| :          :| |  |
Db        395 FITKGAWEVIG-----IEALTTDYLYYISNEYKGMPGGRNLYKI---------QLND--Y 438

Qy        556 SHSCCIS-----QHCDFFISKYSNQKNPH---C----VSLYKLSSPEDDPTCKTKEFWAT 603
              :   |:|     : | ::   :| :   :    |    : || |    :|   :  |   :
Db        439 TKVACLSCELKPERCQYYSVSFSKEAKYYQLRCSGPGLPLYTLHRSSNDEELRVLED-NS 497

Qy        604 ILDSAGPLPDYTPPE---IFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660
              ||    | :   |      |   : | |  | |:  ||      |||| :: :| ||  |
Db        498 ALDKM--LQEVQMPSKKLDFIILNETKF-WYQMILPPH-FDTSKKYPLLIDVYAGPCSQ- 552

Qy        661 VNNRFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLKFEGAFKYKMGQIEIDDQVE 716
                    |    ||||    |||   ::|    | ||| ::| |   |    ::|  |::||:|

Db        553 -----KADAIFRLNWATYLASTENIIVASFDGRGSGYQGDKIMHAVNRRLGTFEVEDQIE 607

Qy        717 GLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 776
                  :   |:  |:|  |: | |||||||:: | |   | :|:  || |||: | :||: ||
Db        608 AARQF-SKMGFVDDKRIAIWGWSYGGYVTSMVLGAGSGVFKCGIAVAPVSRWEYYDSVYT 666

Qy        777 ERYMGHP-DQNEQGYYLGSVAM-QAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVR 834
              |||||  |  |:   ||  |  | :|| |  :    ||:||  |:|||| :: :   ||
Db        667 ERYMGLPTPQDNLDYYKNSTVMSRAENF--KQVEYLLIHGTADDNVHFQQSAQISKALVD 724

Qy        835 AGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQE 870
              ||  :    |  | | |    : :|   |: |::::
Db        725 AGVDFQAMWYTDEDHGIASGPAHQHIYTHMSHFIKQ 760
```

```
RESULT 2
DPP4_BOVIN
ID   DPP4_BOVIN     STANDARD;      PRT;   765 AA.
AC   P81425; Q8WMG8;
DT   15-DEC-1998 (Rel. 37, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE   antigen CD26) (Adenosine deaminase complexing protein) (ADCP-I)
DE   (Activation molecule 3) (ACT3) (WC10).
GN   DPP4 OR CD26.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymphocytes;
RX   MEDLINE=22067734; PubMed=12073152;
RA   Lee S.-U., Park Y.-H., Davis W.C., Hamilton M.J., Naessens J.,
RA   Bohach G.A.;
RT   "Molecular characterization of bovine CD26 upregulated by a
RT   staphylococcal superantigen.";
RL   Immunogenetics 54:216-220(2002).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.
RC   TISSUE=Thymus;
RX   MEDLINE=22021197; PubMed=11981836;
RA   Gliddon D.R., Howard C.J.;
RT   "CD26 is expressed on a restricted subpopulation of dendritic cells in
RT   vivo.";
RL   Eur. J. Immunol. 32:1472-1481(2002).
RN   [3]
RP   SEQUENCE OF 1-24.
RC   TISSUE=T-cell;
RX   MEDLINE=21482004; PubMed=11598101;
RA   Lee S.-U., Ferens W., Davis W.C., Hamilton M.J., Park Y.-H., Fox L.K.,
RA   Naessens J., Bohach G.A.;
RT   "Identity of activation molecule 3 on superantigen-stimulated bovine
RT   cells is CD26.";
RL   Infect. Immun. 69:7190-7193(2001).
RN   [4]
RP   SEQUENCE OF 537-546.
RC   TISSUE=Kidney;
RX   MEDLINE=98293306; PubMed=9629661;
RA   Ben-Shooshan I., Parola A.H.;
RT   "The CP-I subunit of adenosine deaminase complexing protein from calf
RT   kidney is identical to human, mouse, and rat dipeptidyl peptidase
RT   IV.";
RL   Comp. Biochem. Physiol. 119B:289-292(1998).
CC   -!- FUNCTION: Removes N-terminal dipeptides sequentially from
CC       polypeptides having unsubstituted N-termini provided that the
CC       penultimate residue is proline. Binds and regulates the activity
CC       of ADA.
CC   -!- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC       Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC       Xcc is neither Pro nor hydroxyproline.
```

```
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC       a soluble form (By similarity).
CC   -!- TISSUE SPECIFICITY: Intestinal epithelium, dendritic cells and
CC       several immune system tissues.
CC   -!- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC       by proteolytic processing (By similarity).
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF461806; AAL67836.1; -.
DR   EMBL; AY056834; AAL23628.1; -.
DR   MEROPS; S09.003; -.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW   Transmembrane; Glycoprotein; Signal-anchor.
FT   CHAIN         1    765       DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT                                (MDPP).
FT   CHAIN        38    765       DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT                                (SDPP) (BY SIMILARITY).
FT   DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     29       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       30    765       EXTRACELLULAR (POTENTIAL).
FT   ACT_SITE    629    629       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    707    707       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    739    739       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     84     84       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     91     91       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    149    149       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    218    218       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    228    228       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    271    271       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    280    280       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    320    320       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    392    392       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    495    495       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    684    684       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   765 AA;  88369 MW;  E32165421F43E116 CRC64;

  Query Match             12.0%;  Score 566;  DB 1;  Length 765;
  Best Local Similarity   27.0%;  Pred. No. 1.5e-33;
  Matches  191;  Conservative  105;  Mismatches  274;  Indels  138;  Gaps   29;

Qy        206 PADPDWI---------AFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFV 256
              |  :  ||         |:: :|||:: |      :|:|:        ::|    |:  :|
Db        148 PNNTQWITWSSVGHKLAYVWNNDIYVKNEPNSPSQRITWTG------KKDVIYNGITDWV 201

Qy        257 LQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH---VTSPMLETRRADS 312
               :|| |  ||  || | :          : | : :::|| :|     :   |:  :
Db        202 YEEEVFSAYSALWWSPNSTF---------LAYAQFNDTEVPLIEYSFYSDESLQYPKTVK 252

Qy        313 FRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGK 372
              ||| |  ||  | |    : | :    |:   :::: |  :|  |  |:    |  | :
Db        253 IPYPKAGAVNP--TIKFFVVNISSLSPNINATSQQIVPPGSVLI-GDHYLCDVTWVTEER 309

Qy        373 YAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWIN 432
              :                              ::  | |::   |:  :  |: :|  ||:
Db        310 IS----------------------------LQWLRRIQNY--SIMDICDYDRSTGRWIS 338

Qy        433 IHDIFHVFPQSHEEEI------------EFIFASECKTGFRHLYKITSILKESKYKRSSG 480
                  |:       ||            |  | |:  :     ||| |   |  ||

Db        339 SVGRQHI-------EISTTGWVGRFRPAEPHFTSDGNS----FYKIIS--NEEGYKHI-- 383

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGT-KDSPLEHHLYVV 539
                    |:   :    || | |||:|     |:       | |      |  |   :|| :
Db        384 -----CHFQTDKRNCTFITKGAWEVIG-----IEALTSDYLYYISNEYKGMPGARNLYKI 433

Qy        540 SYVNPGEVTRLT-----DRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPT 594
                :  :|| |:      ||   :|   ||   ::  : |     |     || | :  :|
Db        434 QLNDYTKVTCLSCELNPDRCQYYSVSFSQEAKYYQLRCSGPGLP----LYTLHNSNND-- 487

Qy        595 CKTKEFWATILDSAGPLPD-----YTPPEIFSFESTTGFTL-YGMLYKPHDLQPGKKYPT 648
                ||    :|::   |         | :   |    |    | |:  ||     ||||
Db        488 ---KEL--RVLENNSDLDQVLQDVQMPSKKLDFIHLHGTKFWYQMILPPH-FDKSKKYPL 541

Qy        649 VLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLKFEGAFKY 704
              :| :| ||  |       |    ||||    |||   ::|   | ||| ::| |   |
Db        542 LLEVYAGPCSQ------KADAIFRLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINR 595

Qy        705 KMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAP 764
              ::|  |::||:|  :   |:  |:|  |: | ||||||:: | |  | :|:  || ||
Db        596 RLGTFEVEDQIEATRQF-SKMGFVDDKRIAIWGWSYGGYVTSMVLGAGSGVFKCGIAVAP 654

Qy        765 VTLWIFYDTGYTERYMG--HPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHF 822
              |: | :||: |||||||   |: |   |   :|  :|| |  :    ||:||  |:||||
Db        655 VSKWEYYDSVYTERYMGLPTPEDNLDSYRNSTVMSRAENF--KQVEYLLIHGTADDNVHF 712

Qy        823 AHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQE 870
              :: :   || ||  :    |  | | |    : :|   |: |:|::
Db        713 QQSAQISKALVDAGVDFQSMWYTDEDHGIASSTAHQHIYTHMSHFLKQ 760


RESULT 3
DPP4_HUMAN
ID   DPP4_HUMAN     STANDARD;      PRT;   766 AA.
AC   P27487;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE   antigen CD26) (TP103) (Adenosine deaminase complexing protein-2)
DE   (ADABP).
GN   DPP4 OR ADCP2 OR CD26.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92329551; PubMed=1352704;
RA   Misumi Y., Hayashi Y., Arakawa F., Ikehara Y.;
RT   "Molecular cloning and sequence analysis of human dipeptidyl
RT   peptidase IV, a serine proteinase on the cell surface.";
RL   Biochim. Biophys. Acta 1131:333-336(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=95012454; PubMed=7927537;
RA   Abbott C.A., Baker E., Sutherland G.R., McCaughan G.W.;
RT   "Genomic organization, exact localization, and tissue expression of
RT   the human CD26 (dipeptidyl peptidase IV) gene.";
RL   Immunogenetics 40:331-338(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Peripheral blood;
RX   MEDLINE=92325476; PubMed=1352530;
RA   Tanaka T., Camerini D., Seed B., Torimoto Y., Dang N.H., Kameoka J.,
RA   Dahlberg H.N., Schlossman S.F., Morimoto C.;
RT   "Cloning and functional expression of the T cell activation antigen
RT   CD26.";
RL   J. Immunol. 149:481-486(1992).
RN   [4]
RP   ERRATUM.
```

```
RX   MEDLINE=93171637; PubMed=8094732;
RA   Tanaka T.;
RL   J. Immunol. 150:2090-2090(1993).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Prostate;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [6]
RP   SEQUENCE OF 1-551 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=92165847; PubMed=1347043;
RA   Darmoul D., Lacasa M., Baricault L., Marguet D., Sapin C.,
RA   Trotot P., Barbat A.;
RT   "Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like
RT   colon cancer cell lines HT-29 and Caco-2. Cloning of the complete
RT   human coding sequence and changes of dipeptidyl peptidase IV mRNA
RT   levels during cell differentiation.";
RL   J. Biol. Chem. 267:4824-4833(1992).
RN   [7]
RP   SEQUENCE OF 545-766 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=91024044; PubMed=1977364;
RA   Darmoul D., Lacasa M., Chantret I., Swallow D., Trugnan G.;
RT   "Isolation of a cDNA probe for the human intestinal
RT   dipeptidylpeptidase IV and assignment of the gene locus DPP4 to
RT   chromosome 2.";
RL   Ann. Hum. Genet. 54:191-197(1990).
RN   [8]
RP   SEQUENCE OF 1-31 FROM N.A.
RX   MEDLINE=96067599; PubMed=7487939;
RA   Boehm S.K., Gum J.R. Jr., Erickson R.H., Hicks J.W., Kim Y.S.;
RT   "Human dipeptidyl peptidase IV gene promoter: tissue-specific
RT   regulation from a TATA-less GC-rich sequence characteristic of a
RT   housekeeping gene promoter.";
RL   Biochem. J. 311:835-843(1995).
RN   [9]
RP   PARTIAL SEQUENCE.
RC   TISSUE=Kidney;
RX   MEDLINE=93210468; PubMed=8096237;
RA   Morrison M.E., Vijayasaradhi S., Engelstein D., Albino A.P.,
RA   Houghton A.N.;
RT   "A marker for neoplastic progression of human melanocytes is a cell
RT   surface ectopeptidase.";
RL   J. Exp. Med. 177:1135-1143(1993).
CC   -!- FUNCTION: Removes N-terminal dipeptides sequentially from
CC       polypeptides having unsubstituted N-termini provided that the
CC       penultimate residue is proline. Plays a role in T cell activation.
CC   -!- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC       Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC       Xcc is neither Pro nor hydroxyproline.
CC   -!- SUBUNIT: Homodimer or heterodimer with Seprase (FAP).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC       a soluble form.
CC   -!- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC       by proteolytic processing.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD26 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd26.htm".
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U13735; AAB60646.1; -.
DR   EMBL; U13710; AAB60646.1; JOINED.
DR   EMBL; U13711; AAB60646.1; JOINED.
DR   EMBL; U13712; AAB60646.1; JOINED.
DR   EMBL; U13713; AAB60646.1; JOINED.
DR   EMBL; U13714; AAB60646.1; JOINED.
DR   EMBL; U13715; AAB60646.1; JOINED.
DR   EMBL; U13716; AAB60646.1; JOINED.
DR   EMBL; U13717; AAB60646.1; JOINED.
DR   EMBL; U13718; AAB60646.1; JOINED.
DR   EMBL; U13719; AAB60646.1; JOINED.
DR   EMBL; U13720; AAB60646.1; JOINED.
DR   EMBL; U13721; AAB60646.1; JOINED.
DR   EMBL; U13722; AAB60646.1; JOINED.
DR   EMBL; U13723; AAB60646.1; JOINED.
DR   EMBL; U13724; AAB60646.1; JOINED.
DR   EMBL; U13725; AAB60646.1; JOINED.
DR   EMBL; U13726; AAB60646.1; JOINED.
DR   EMBL; U13727; AAB60646.1; JOINED.
DR   EMBL; U13728; AAB60646.1; JOINED.
DR   EMBL; U13729; AAB60646.1; JOINED.
DR   EMBL; U13730; AAB60646.1; JOINED.
DR   EMBL; U13731; AAB60646.1; JOINED.
DR   EMBL; U13732; AAB60646.1; JOINED.
DR   EMBL; U13733; AAB60646.1; JOINED.
DR   EMBL; U13734; AAB60646.1; JOINED.
DR   EMBL; M74777; AAA51943.1; -.
DR   EMBL; BC013329; AAH13329.1; -.
DR   EMBL; M80536; AAA52308.1; -.
DR   EMBL; X60708; CAA43118.1; -.
DR   EMBL; S79876; AAB35614.1; -.
DR   PIR; S24313; CDHU26.
DR   PDB; 1N1M; 07-JAN-03.
DR   MEROPS; S09.003; -.
DR   Genew; HGNC:3009; DPP4.
DR   MIM; 102720; -.
DR   GO; GO:0004274; F:dipeptidyl-peptidase IV activity; TAS.
DR   GO; GO:0005209; F:plasma protein; TAS.
DR   GO; GO:0006955; P:immune response; TAS.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW   Transmembrane; Glycoprotein; Signal-anchor; 3D-structure.
FT   CHAIN         1    766       DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT                                (MDPP).
FT   CHAIN        39    766       DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT                                (SDPP).
FT   DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     28       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       29    766       EXTRACELLULAR (POTENTIAL).
FT   ACT_SITE    630    630       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    708    708       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    740    740       CHARGE RELAY SYSTEM (BY SIMILARITY).
```

```
FT   CARBOHYD     85     85       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     92     92       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    150    150       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    219    219       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    229    229       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    281    281       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    321    321       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    520    520       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    685    685       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      6      6       K -> R (IN REF. 6).
FT   CONFLICT      7      7       V -> I (IN REF. 1).
FT   CONFLICT    437    437       S -> I (IN REF. 1).
FT   CONFLICT    557    557       T -> I (IN REF. 7).
FT   CONFLICT    663    663       D -> E (IN REF. 7).

  Query Match           11.2%;  Score 525;  DB 1;  Length 766;
  Best Local Similarity  26.3%;  Pred. No. 1.5e-30;
  Matches  186;  Conservative  108;  Mismatches  277;  Indels  136;  Gaps   33;

Qy        206 PADPPDWI---------AFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFV 256
              |  :  |:          |:: :|||::          |:|:          :||    |:  :|
Db        149 PNNTQWVTWSPVGHKLAYVWNNDIYVKIEPNLPSYRITWTG------KEDIIYNGITDWV 202

Qy        257 LQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH---VTSPMLETRRADS 312
               :|| |  ||  || |          : | : :::|| :|     :   |:  :
Db        203 YEEEVFSAYSALWWSPNGTF---------LAYAQFNDTEVPLIEYSFYSDESLQYPKTVR 253

Qy        313 FRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGK 372
               |||  |  || | |  :  :  |:   : :    ::  |  :|  |  |:    |
Db        254 VPYPKAGAVNPTVKFFV--VNTDSLSSVTNATSIQITAPASMLI-GDHYLCDVTWA---- 306

Qy        373 YAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWIN 432
                    :| |: :  :                  | |::   ||  :  |:|::  | |
Db        307 ---------TQERISLQWL---------------RRIQNY--SVMDICDYDESSGRW-N 338

Qy        433 IHDIFHVFPQSHEEEIEFIFASECKTGFRH-------LYKITSILKESKYKRSSGGLPAP 485
                    :  : | |     :    :    |          |||  |   |  |:
Db        339 C-----LVARQHIEMSTTGWVGRFRPSEPHFTLDGNSFYKIIS--NEEGYRHI------- 384

Qy        486 SDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGT-KDSPLEHHLYVVSYVNP 544
               |:   |:   || | |||:|     |:       | |      |  |   :|| :
Db        385 CYFQIDKKDCTFITKGTWEVIG-----IEALTSDYLYYISNEYKGMPGGRNLYKI----- 434

Qy        545 GEVTRLTDRGYSHSCCIS-----QHCDFFISKYSNQKNPH---C----VSLYKLSSPEDD 592
                  :|:|  |:   |:|     : | ::   :| :   :    |    : || | |  :|
Db        435 ----QLSD--YTKVTCLSCELNPERCQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVND 488

Qy        593 PTCKTKEFWATILDSAGPLPDYTPPE---IFSFESTTGFTLYGMLYKPHDLQPGKKYPTV 649
                  : |   : ||   | :   |     |  : | |  | |:  ||      |||| :
Db        489 KGLRVLED-NSALDKM--LQNVQMPSKKLDFIILNETKF-WYQMILPPH-FDKSKKYPLL 543

Qy        650 LFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLKFEGAFKYK 705
               | :| ||  |      |    ||||    |||   ::|   | ||| ::| |    |   :
Db        544 LDVYAGPCSQ------KADTVFRLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRR 597

Qy        706 MGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPV 765
               :|  |::||:|  :   |:  |:|  |: | ||||||:: | |   | :|:  || |||
Db        598 LGTFEVEDQIEAARQF-SKMGFVDNKRIAIWGWSYGGYVTSMVLGSGSGVFKCGIAVAPV 656

Qy        766 TLWIFYDTGYTERYMG--HPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFA 823
               : | :||: ||||||    |: |   |   :|  :|| |  :    ||:||  |:||||
Db        657 SRWEYYDSVYTERYMGLPTPEDNLDHYRNSTVMSRAENF--KQVEYLLIHGTADDNVHFQ 714

Qy        824 HTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLLHYLQE 870
               :: :   ||  |  :    |  | | |    : :|   |: |::::
Db        715 QSAQISKALVDVGVDFQAMWYTDEDHGIASSTAHQHIYTHMSHFIKQ 761


RESULT 4
DPP4_RAT
ID   DPP4_RAT       STANDARD;      PRT;   767 AA.
AC   P14740;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE   antigen CD26) (GP110 glycoprotein) (Bile canaliculus domain-specific
DE   membrane glycoprotein).
GN   DPP4 OR CD26.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89123496; PubMed=2563382;
RA   Ogata S., Misumi Y., Ikehara Y.;
RT   "Primary structure of rat liver dipeptidyl peptidase IV deduced from
RT   its cDNA and identification of the NH2-terminal signal sequence as
RT   the membrane-anchoring domain.";
RL   J. Biol. Chem. 264:3596-3601(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88068516; PubMed=3479775;
RA   Hong W., Doyle D.;
RT   "cDNA cloning for a bile canaliculus domain-specific membrane
RT   glycoprotein of rat hepatocytes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:7962-7966(1987).
RN   [3]
RP   SIGNAL-ANCHOR.
RX   MEDLINE=90338089; PubMed=1974258;
RA   Hong W., Doyle D.;
RT   "Molecular dissection of the NH2-terminal signal/anchor sequence of
RT   rat dipeptidyl peptidase IV.";
RL   J. Cell Biol. 111:323-328(1990).
RN   [4]
RP   SEQUENCE OF 281-302.
RC   TISSUE=Kidney;
RX   MEDLINE=94128239; PubMed=7905271;
RA   Iwaki-Egawa S., Watanabe Y., Fujimoto Y.;
RT   "N-terminal amino acid sequence of the 60-kDa protein of rat kidney
RT   dipeptidyl peptidase IV.";
RL   Biol. Chem. Hoppe-Seyler 374:973-975(1993).
CC   -!- FUNCTION: REMOVES N-TERMINAL DIPEPTIDES SEQUENTIALLY FROM
CC       POLYPEPTIDES HAVING UNSUBSTITUTED N-TERMINI PROVIDED THAT THE
CC       PENULTIMATE RESIDUE IS PROLINE.
CC   -!- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC       Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC       Xcc is neither Pro nor hydroxyproline.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS IN
CC       A SOLUBLE FORM.
CC   -!- PTM: THE SOLUBLE FORM (SDPP) DERIVES FROM THE MEMBRANE FORM (MDPP)
CC       BY PROTEOLYTIC PROCESSING.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04591; AAA41096.1; -.
DR   EMBL; J02997; AAA41272.1; -.
DR   PIR; A39914; A39914.
DR   MEROPS; S09.003; -.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
```

```
KW   Transmembrane; Glycoprotein; Signal-anchor.
FT   CHAIN         1    767       DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT                                (MDPP).
FT   CHAIN        37    767       DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT                                (SDPP).
FT   CHAIN       281    767       DIPEPTIDYL PEPTIDASE IV 60 KDA SOLUBLE
FT                                FORM.
FT   DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     28       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       29    767       EXTRACELLULAR (POTENTIAL).
FT   ACT_SITE    631    631       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    709    709       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    741    741       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     83     83       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     90     90       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    148    148       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    217    217       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    227    227       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    319    319       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    521    521       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    686    686       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     38     38       A -> R (IN REF. 2).
FT   CONFLICT    183    183       I -> T (IN REF. 2).
FT   CONFLICT    332    332       T -> N (IN REF. 2).
FT   CONFLICT    352    352       C -> V (IN REF. 2).
FT   CONFLICT    394    394       V -> D (IN REF. 2).
FT   CONFLICT    562    562       L -> F (IN REF. 2).
FT   CONFLICT    624    624       R -> Q (IN REF. 2).
SQ   SEQUENCE   767 AA;  88003 MW;  AA17C0EC6F0E4652 CRC64;

  Query Match             11.1%;  Score 523.5;  DB 1;  Length 767;
  Best Local Similarity   26.0%;  Pred. No. 2e-30;
  Matches  188;  Conservative  104;  Mismatches  266;  Indels  165;  Gaps   33;

Qy        206 PADPDWI---------AFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFV 256
              |  :  ||         |::  |||::         |:|      :  :|:    |:  :|
Db        147 PNNTQWITWSQEGHKLAYVWKNDIYVKIEPHLPSHRIT------STGKENVIFNGINDWV 200

Qy        257 LQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRY 315
              :|| |  ||  ||      :|:|  :    : :     :|    :   |:  :      |
Db        201 YEEEIFGAYSALWW------SPNGTFLAYAQFNDTGVPLIEYSFYSDESLQYPKTVWIPY 254

Qy        316 PKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAW 375
              || |  || | |  :  :  |:       |  ::  |  :   |  |:    |  | : :
Db        255 PKAGAVNPTVKFFI--VNTDSLSSTTTTIPMQITAPASVT-TGDHYLCDVAWVSEDRIS- 310

Qy        376 SILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHD 435
                                          ::  | |::   ||  :  |::|| :| |
Db        311 ----------------------------LQWLRRIQNY--SVMAICDYDKTTLVW-NC-- 337

Qy        436 IFHVFPQSHEEEIEFIFASEC----------------------KTGFRHLYKITSILKE 472
                  : :| ||       |                      | |::|   |    |:
Db        338 ------PTTQEHIETSATGWCGRFRPAEPHFTSDGSSFYKIVSDKDGYKH---ICQFQKD 388

Qy        473 SKYKRSSGGLPAPSDFKCPIKEEIA--ITSGEWEVLGRHGSNIQVDEVRRLVYFEGT-KD 529
                |                    |::   || | |||:     :|:      | |     |:
Db        389 RK-----------------PEQVCTFITKGAWEVI-----SIEALTSDYLYYISNEYKE 425

Qy        530 SPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCIS-----QHCDFF---ISKYSNQKNPHC- 580
                |   :|| :          :|||  :::  |:|     : | ::   :|| :      |
Db        426 MPGGRNLYKI---------QLTD--HTNKKCLSCDLNPERCQYYSVSLSKEAKYYQLGCR 474

Qy        581 ---VSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPE---IFSFESTTGFTLYGML 634
                : || |    |    :  |   : ||    | |  |      |   : | |  | |:
Db        475 GPGLPLYTLHRSTDQKELRVLED-NSALDKM--LQDVQMPSKKLDFIVLNETRF-WYQMI 530

Qy        635 YKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN---TLASLGYVVVV-IDNRG 690
                ||     |||| :: :| ||  |       |    ||||    |||   ::|   | ||
Db        531 LPPH-FDKSKKYPLLIDVYAGPCSQ------KADAAFRLNWATYLASTENIIVASFDGRG 583

Qy        691 SCHRGLKFEGAFKYKMGQIEIDDQVEGL-QYLASRYDFIDLDRVGIHGWSYGGYLSLMAL 749
              | ::| |   |   ::| :|::||:|   |:|  :  |:|  || | |||||||:: | |

Db        584 SGYQGDKIMHAINKRLGTLEVEDQIEAARQFL--KMGFVDSKRVAIWGWSYGGYVTSMVL 641

Qy        750 MQRSDIFRVAIAGAPVTLWIFYDTGYTERYMG--HPDQNEQGYYLGSVAMQAEKFPSEPN 807
                  | :|:  || |||: | :||: |||||||   |: |   |   :|  :|| |  :
Db        642 GSGSGVFKCGIAVAPVSRWEYYDSVYTERYMGLPTPEDNLDHYRNSTVMSRAENF--KQV 699

Qy        808 RLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHY 867
                ||:||  |:|||| :: :   || ||  :   |  | | |   : :|   |: |:
Db        700 EYLLIHGTADDNVHFQQSAQISKALVDAGVDFQAMWYTDEDHGIASSTAHQHIYSHMSHF 759

Qy        868 LQE 870
                ||:
Db        760 LQQ 762


RESULT 5
YEA8_SCHPO
ID   YEA8_SCHPO     STANDARD;      PRT;   793 AA.
AC   O14073;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Putative dipeptidyl aminopeptidase C2E11.08 (EC 3.4.14.-).
GN   SPACUNK4.08 OR SPAC2E11.08.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like
CC       vacuoles (By similarity).
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL031180; CAA20138.1; -.
DR   PIR; T41703; T41703.
```

```
DR   GeneDB_SPombe; SPACUNK4.08; -.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hypothetical protein; Hydrolase; Aminopeptidase; Dipeptidase;
KW   Serine protease; Transmembrane; Glycoprotein; Signal-anchor.
FT   DOMAIN        1     24       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     25     45       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       46    793       LUMENAL (POTENTIAL).
FT   ACT_SITE    647    647       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    722    722       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    755    755       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD    101    101       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    136    136       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    246    246       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    299    299       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    303    303       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    324    324       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    336    336       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    377    377       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    384    384       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    407    407       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    535    535       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    761    761       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   793 AA;  91304 MW;  20B707F97F231463 CRC64;

  Query Match             11.1%;  Score 519.5;  DB 1;  Length 793;
  Best Local Similarity   26.3%;  Pred. No. 4e-30;
  Matches  187;  Conservative  113;  Mismatches  256;  Indels  155;  Gaps   32;

Qy        212 IAFIHSNDIWI----SNIVTREERRLTY-----VHNELANMEEDARSAGVATFVLQEE-F 261
              ::|:::||:::     |:     :||||     | |             |:  :: :||
Db        183 LSFVYNNDLYVRKNDGNV-----QRLTYDGTVDVFN-----------GLTDWIYEEEVL 225

Qy        262 DRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV---------TSPMLETRRADS 312
              |  || | ::          :| : : :|||:   |          : |  :  :  :
Db        226 SSPSTIWWSPDSD---------KIAFLKLNESEIPTYHYPLYTAELDPSLPEFDYNKDMA 276

Qy        313 FRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGK 372
             :||| |  || |:  :::: :|       :: | :|   : : | ::  :
Db        277 IKYPKPGNPNPSVSLFVADLNSNASSNFSLWHNEPLAEP---VVQNVLWVNTS------- 326

Qy        373 YAWSILL---DRSQTRLQIVLISPEL--FIPVEDDVMER-----QRLIESVPDSVTPLII 422
              |:|:   :|: | :   |:  ||     |: : :|      |:  :  |  :   ::
Db        327 ---SVLVQFTNRNSTCITARLLDTELKSIHTVKTECLEEGWYEVQQSAKMFP--LNNSLV 381

Qy        423 YEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGL 482
              :|  :|      ||  :   :|   | |                            :|
Db        382 WENWSD---GYFDILALDDYNHLAFIPF---------------------------NGSS 410

Qy        483 PAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYV 542
              |          | :||| |:|       | :|     |||  |     | ||| || :
Db        411 P-----------IYLTSGAWDVT---DGPIHIDGDFGNVYFLATLKDSTERHLYYVS-L 454

Qy        543 NPGEVTRLTDRGYS---HSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKE 599
              :  |:  :|| |      :|   |   ||::  |     |      :| | :|   | : |
Db        455 DTLEIYGITDNGEDEGYYSTSFSPFGDFYVLNYHGPDVP----WQELRSTKDKDYCLSLE 510

Qy        600 FWATILDSAGPLPDYTPPEI----FSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGG 655
              : :      |  | | :     :|  || |     | :| :    |||| : | |||
Db        511 TNSRLKQQ---LSSITLPSVEYGKLTFNDTT-FNF--MERRPRNFDVNKKYPVLFFAYGG 564

Qy        656 PQVQLVNNRFKGVKYFRLNTLASLG------YVVVVIDNRGSCHRGLKFEGAFKYKMGQI 709
              |  | |        | ||::  | |       ::|| :| ||:   |  |  :    :|:
Db        565 PGSQQV------AKLFRVDFQAYLASHPDFEFIVVTLDGRGTGFNGNAFRYSVSRHLGEW 618

Qy        710 EIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWI 769
              |  || :  :: |    |:| : ||| ||||||||:|  | :  |:|   :| |||| |

Db        619 ESYDQGQAGKFWAD-LPFVDENHVGIWGWSYGGYLTLKTL-ETQDVFSYGMAVAPVTDWR 676

Qy        770 FYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILL 829
              ||: ||||||  |  |::| |  |      |||  :  |  : ||  |:|||| |:  |:
Db        677 LYDSVYTERYMDLPQYNKEG-YKNSQIHDYEKF-KQLKRFFVAHGTGDDNVHFQHSMHLM 734

Qy        830 SFLVRAG-KPYDLQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAAL 879
              |  |    ||: ::|   |||    |  :  | : | | | :|  :  :
Db        735 DGLNLANCYNYDMAVFPDSAHSI----SYHNASLSIYHRLSEWIGDALGRI 781


RESULT 6
DPP4_MOUSE
ID   DPP4_MOUSE     STANDARD;      PRT;   760 AA.
AC   P28843;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE   antigen CD26) (Thymocyte-activating molecule) (THAM).
GN   DPP4 OR CD26.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Swiss; TISSUE=Thymus;
RX   MEDLINE=92129288; PubMed=1370813;
RA   Marguet D.A., Bernard A.-M., Vivier I., Darmoul D., Naquet P.,
RA   Pierres M.;
RT   "cDNA cloning for mouse thymocyte-activating molecule. A
RT   multifunctional ecto-dipeptidyl peptidase IV (CD26) included in a
RT   subgroup of serine proteases.";
RL   J. Biol. Chem. 267:2200-2208(1992).
RN   [2]
RP   REVISIONS.
RA   Marguet D.A.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B10.A; TISSUE=Liver;
RX   MEDLINE=95092780; PubMed=7999781;
RA   Bernard A.-M., Mattei M.-G., Pierres M., Marguet D.;
RT   "Structure of the mouse dipeptidyl peptidase IV (CD26) gene.";
RL   Biochemistry 33:15204-15214(1994).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SEQUENCE OF 1-20.
RX   MEDLINE=91302787; PubMed=1712807;
RA   Vivier I., Marguet D.A., Naquet P., Bonicel J., Black D., Li C.X.-Y.,
```

```
RA   Bernard A.-M., Gorvel J.-P., Pierres M.;
RT   "Evidence that thymocyte-activating molecule is mouse CD26
RT   (dipeptidyl peptidase IV).";
RL   J. Immunol. 147:447-454(1991).
CC   -!- FUNCTION: Removes N-terminal dipeptides sequentially from
CC       polypeptides having unsubstituted N-termini provided that the
CC       penultimate residue is proline.
CC   -!- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC       Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC       Xcc is neither Pro nor hydroxyproline.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC       a soluble form.
CC   -!- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
CC       by proteolytic processing.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X58384; CAA41274.1; -.
DR   EMBL; U12620; AAA82213.1; -.
DR   EMBL; U12599; AAA82213.1; JOINED.
DR   EMBL; U12600; AAA82213.1; JOINED.
DR   EMBL; U12601; AAA82213.1; JOINED.
DR   EMBL; U12602; AAA82213.1; JOINED.
DR   EMBL; U12603; AAA82213.1; JOINED.
DR   EMBL; U12604; AAA82213.1; JOINED.
DR   EMBL; U12605; AAA82213.1; JOINED.
DR   EMBL; U12606; AAA82213.1; JOINED.
DR   EMBL; U12607; AAA82213.1; JOINED.
DR   EMBL; U12608; AAA82213.1; JOINED.
DR   EMBL; U12609; AAA82213.1; JOINED.
DR   EMBL; U12610; AAA82213.1; JOINED.
DR   EMBL; U12611; AAA82213.1; JOINED.
DR   EMBL; U12612; AAA82213.1; JOINED.
DR   EMBL; U12613; AAA82213.1; JOINED.
DR   EMBL; U12614; AAA82213.1; JOINED.
DR   EMBL; U12615; AAA82213.1; JOINED.
DR   EMBL; U12616; AAA82213.1; JOINED.
DR   EMBL; U12617; AAA82213.1; JOINED.
DR   EMBL; U12618; AAA82213.1; JOINED.
DR   EMBL; U12619; AAA82213.1; JOINED.
DR   EMBL; BC022183; AAH22183.1; -.
DR   MEROPS; S09.003; -.
DR   MGD; MGI:94919; Dpp4.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW   Transmembrane; Glycoprotein; Signal-anchor.
FT   CHAIN         1    760       DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT                                (MDPP).
FT   CHAIN        37    760       DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT                                (SDPP).
FT   DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     28       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       29    760       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      295    466       CYS-RICH.
FT   ACT_SITE    624    624       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    702    702       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    734    734       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     83     83       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     90     90       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    113    113       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    213    213       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    223    223       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    315    315       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    328    328       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    514    514       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    679    679       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   760 AA;  87436 MW;  A5F644B46E4A3DF8 CRC64;

  Query Match             11.0%;  Score 517.5;  DB 1;  Length 760;
  Best Local Similarity   25.0%;  Pred. No. 5.3e-30;
  Matches  203;  Conservative  107;  Mismatches  306;  Indels  197;  Gaps   37;

Qy        131 MYSREEELLRERKRIGTVGI-------ASYDYHQGSGTFLF--------------QAGS 168
              :| :|  :|       |  |         |: ||  |   ||              |
Db         67 LYKQENNILLLNAEHGNSSIFLENSTFESFGYHSVSPDRLFVLLEYNYVKQWRHSYTASY 126

Qy        169 GIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWI---------AFIHSND 219
              || |          |  : |   ||         :  ||         |::  ||
Db        127 NIYDVNK-----------RQLITEEKIPN----------NTQWITWSPEGHKLAYVWKND 165

Qy        220 IWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEE-FDRYSGYWWCPK------ 272
              |::          |:|       :  ||:    |:  :| :|| |  ||  || |
Db        166 IYVKVEPHLPSHRIT------STGEENVIYNGITDWVYEEEVFGAYSALWWSPNNTFLAY 219

Qy        273 AETTPSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEI 332
              |:   :|  ::   :  :|||            |:  :     ||| |  || | | :  :
Db        220 AQFNDTGVPLIEYSF-YSDES----------LQYPKTVWIPYPKAGAVNPTVKFFI--V 265

Qy        333 MIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYA---------WSILLDRSQ 383
              ||:          ::  |  :   |  |:    |  | : :          :|::
Db        266 NIDSLSSSSSAAPIQIPAPASVA-RGDHYLCDVVWATEERISLQWLRRIQNYSVMAICDY 324

Qy        384 TRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQS 443
              :: :   |           |:| :              | :|  |:          |
Db        325 DKINLTWNCPS----------EQQHV-------------EMSTTGWVG--RFRPAEPHF 358

Qy        444 HEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFK-CPIKEEIAITSGE 502
              :   |       | |::|:                        | | | |       || |
Db        359 TSDGSSFYKIISDKDGYKHICHF-----------------PKDKKDCTF-----ITKGA 395

Qy        503 WEVLGRHGSNIQVDEVRRLVYFEGT-KDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCI 561
              |||:      :|:      | |     |: |  :|| :           :|||  :::  |:
Db        396 WEVI-----SIEALTSDYLYYISNQYKEMPGGRNLYKI---------QLTD--HTNVKCL 439

Qy        562 S-----QHCDFFISKYSNQKNPH---C----VSLYKLSSPEDDPTCKTKEFWATILDSAG 609
              |     : | ::   :| :   :   |    : || |     |   :  |       |::
Db        440 SCDLNPERCQYYAVSFSKEAKYYQLGCWGPGLPLYTLHRSTDHKELRVLE------DNSA 493

Qy        610 ---PLPDYTPPE---IFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 663
                | |   |     |  : | |  | |:  ||     |||| :| :| || |
Db        494 LDRMLQDVQMPSKKLDFIVLNETRF-WYQMILPPH-FDKSKKYPLLLDVYAGPCSQ---- 547

Qy        664 RFKGVKYFRLN---TLASLGYVVVV-IDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQ 719
              |    ||||    |||   ::|   | ||| ::| |  |    |   ::| :|::||:|  :
Db        548 --KADASFRLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRRLGTLEVEDQIEAAR 605

Qy        720 YLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERY 779
              :  |:|  || | |||||||:: | |   | :|:  || |||: | :||: |||||
Db        606 QFV-KMGFVDSKRVAIWGWSYGGYVTSMVLGSGSGVFKCGIAVAPVSRWEYYDSVYTERY 664

Qy        780 MG--HPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGK 837
              ||   |: |   |   :|  :|| |  :    ||:||  |:|||  :: :   || ||
Db        665 MGLPIPEDNLDHYRNSTVMSRAEHF--KQVEYLLIHGTADDNVHFQQSAQISKALVDAGV 722

Qy        838 PYDLQIYPQERHSIRVPESGEHYELHLLLHYLQE 870
              :   |  | | |    : :|   |: |:||:
Db        723 DFQAMWYTDEDHGIASSTAHQHIYSHMSHFLQQ 755


RESULT 7
DAP2_YEAST
```

```
ID   DAP2_YEAST     STANDARD;      PRT;   818 AA.
AC   P18962;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Dipeptidyl aminopeptidase B (EC 3.4.14.-) (DPAP B) (YSCV).
GN   DAP2 OR YHR028C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89174971; PubMed=2647766;
RA   Roberts C.J., Pohlig G., Rothman J.H., Stevens T.H.;
RT   "Structure, biosynthesis, and localization of dipeptidyl
RT   aminopeptidase B, an integral membrane glycoprotein of the yeast
RT   vacuole.";
RL   J. Cell Biol. 108:1363-1373(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=94378003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
CC       VACUOLES.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X15484; CAA33512.1; -.
DR   EMBL; U10399; AAB68879.1; -.
DR   PIR; S46780; A30107.
DR   MEROPS; S09.006; -.
DR   SGD; S0001070; DAP2.
DR   GO; GO:0000329; C:vacuolar membrane (sensu Fungi); IDA.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW   Transmembrane; Glycoprotein; Signal-anchor.
FT   DOMAIN        1     29       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     30     45       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       46    818       LUMENAL (POTENTIAL).
FT   ACT_SITE    679    679       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    756    756       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    789    789       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     63     63       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     79     79       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    139    139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    372    372       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    392    392       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    421    421       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    738    738       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     83     83       Q -> H (IN REF. 1).
FT   CONFLICT    125    125       S -> N (IN REF. 1).
FT   CONFLICT    182    188       FEEIGNE -> LRRLET (IN REF. 1).
FT   CONFLICT    200    200       D -> N (IN REF. 1).
FT   CONFLICT    366    375       TSNVVRNESS -> DFKRGKERKF (IN REF. 1).
FT   CONFLICT    808    818       AKRAFDGQFVK -> QSVLSMGNLTNELTIYSSSHRDIHKT
FT                                FSYLHTMYI (IN REF. 1).
SQ   SEQUENCE   818 AA;  93404 MW;  318F450445375BD3 CRC64;

  Query Match             10.3%;  Score 484.5;  DB 1;  Length 818;
  Best Local Similarity   26.0%;  Pred. No. 1.6e-27;
  Matches  183;  Conservative   95;  Mismatches  250;  Indels  177;  Gaps   35;

Qy        212 IAFIHSNDIWISNIVTREERRLTYVHNE----LANMEEDARSAGVATFVLQEE-FDRYSG 266
              ||::  |:|:| : ::::  |   | |:    | | : |          :| :|| |:
Db        201 IAYVQDNNIYIYSAISKKTIRA--VTNDGSSFLFNGKPD--------WVYEEEVFEDDKA 250

Qy        267 YWWCPKAETTPSGGKILRILYEENDESEV-EIIHVTSPMLETRRAD------SFRYPKTG 319
               || |  :         : : : ||||| | |    |      |       | :|||:|
Db        251 AWWSPTGD---------YLAFLKIDESEVGEFI---IPYYVQDEKDIYPEMRSIKYPKSG 298

Qy        320 TANPKVTFKMSEIMIDAEGRIIDVIDKELIQP--FEILFEGVEYIARAGWTPEGKYAWSI 377
              | ||         || : : |    |       :|  |   |  |   ::
Db        299 TPNP-----------HAELWVYSMKDGTSFHPRISGNKKDGSLLITEVTWVGNG----NV 343

Qy        378 LL---DRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINI- 433
              |:   |||  | : ||         | : :  :: :          | :  | |
Db        344 LVKTTDRSSDILTVFLI---------DTIAKTSNVVRN-----------ESSNGGWWEIT 383

Qy        434 HDIFHV-----FPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDF 488
              |:  :    | : |  :: :        |: ||          | :|        | :
Db        384 HNTLFIPANETFDRPHNGYVDILPIG----GYNHL----------AYFENSNS----SHY 425

Qy        489 KCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVT 548
              |       :| |:|||:      : | :   :|| |: |  | |:| :  :| |:
Db        426 K-------TLTEGKWEVV---NGPLAFDSMENRLYFISTRKSSTERHVYYIDLRSPNEII 475

Qy        549 RLTDRGYS--HSCCISQHCDFFISKYSNQKNPH-------------C-------VSLYKL 586
               :||     :    |    | :  |    | |:                |       ||| |
Db        476 EVTDTSEDGVYDVSFSSGRRFGLLTYKGPKVPYQKIVDFHSRKAEKCDKGNVLGKSLYHL 535

Qy        587 SSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYG--MLYKPHDLQP-- 642
                   |         :|    | ||  |       |    |   :|  :|   ::: |
Db        536 EKNE-------------VLTKI--LEDYAVPRKSFRELNLGKDEFGKDILVNSYEILPND 580

Qy        643 -----GKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLN----TLASLGYVVVVIDNRGSCH 693
                     ||   | ||||  | |          || | :        : |  :|||:| ||:
Db        581 FDETLSDHYPVFFFAYGGPNSQQV------VKTFSVGFNEVVASQLNAIVVVVDGRGTGF 634

Qy        694 RGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRS 753
              :|  |    : ::|   |  ||:          |   |:|  :: : ||||||||:| | :
Db        635 KGQDFRSLVRDRLGDYEARDQISAASLYGS-LTFVDPQKISLFGWSYGGYLTLKTLEKDG 693

Qy        754 D-IFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSV----AMQAEKFPSEPNR 808
                |:  :: |||| | |||: ||||||  | :|  ||   ||   |:        :: ||
Db        694 GRHFKYGMSVAPVTDWRFYDSVYTERYMHTPQENFDGYVESSVHNVTAL------AQANR 747

Qy        809 LLLLHGFLDENVHFAHTSILLSFLVRAG-KPYDLQIYPQERHSIR 852
                ||:||  |:||||  ::     |    |   | : ||: ::|     ||||
Db        748 FLLMHGTGDDNVHFQNSLKFLDLLDLNGVENYDVHVFPDSDHSIR 792


RESULT 8
SEPR_MOUSE
ID   SEPR_MOUSE     STANDARD;      PRT;   761 AA.
AC   P97321;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Seprase (EC 3.4.21.-) (Fibroblast activation protein alpha) (Integral
DE   membrane serine protease).
```

```
GN   FAP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC   STRAIN=BALB/c; TISSUE=Embryo;
RX   MEDLINE=97284459; PubMed=9139873;
RA   Niedermeyer J., Scanlan M.J., Garin-Chesa P., Daiber C., Fiebig H.H.,
RA   Old L.J., Rettig W.J., Schnapp A.;
RT   "Mouse fibroblast activation protein: molecular cloning, alternative
RT   splicing and expression in the reactive stroma of epithelial
RT   cancer.";
RL   Int. J. Cancer 71:383-389(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Breast;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: May have a role in tissue remodeling during development
CC       and wound healing, and contribute to invasiveness in malignant
CC       cancers.
CC   -!- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and
CC       type IV collagen, but not native type I or type IV collagen. Does
CC       not cleave laminin, fibronectin, fibrin or casein.
CC   -!- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is
CC       inactive (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=1;
CC         IsoId=P97321-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P97321-2; Sequence=VSP_005368;
CC       Name=3;
CC         IsoId=P97321-3; Sequence=VSP_005369;
CC   -!- TISSUE SPECIFICITY: Detected in fibroblasts, in placenta, uterus,
CC       embryos from day 7-19 and in new-born mice (P1).
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y10007; CAA71116.1; -.
DR   EMBL; BC019190; AAH19190.1; -.
DR   MEROPS; S09.007; -.
DR   MGD; MGI:109608; Fap.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Protease; Serine protease; Transmembrane; Signal-anchor;
KW   Glycoprotein; Alternative splicing.
FT   DOMAIN        1      4       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      5     25       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       26    761       EXTRACELLULAR (POTENTIAL).
FT   ACT_SITE    624    624       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    702    702       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    734    734       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     49     49       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     92     92       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     99     99       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    227    227       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    314    314       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    679    679       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC     31     35       Missing (in isoform 2).
FT                                /FTId=VSP_005368.
FT   VARSPLIC     31     63       Missing (in isoform 3).
FT                                /FTId=VSP_005369.
FT   CONFLICT    737    737       S -> L (IN REF. 2).
SQ   SEQUENCE   761 AA;  87944 MW;  9174C3AEDA213B25 CRC64;

  Query Match             10.2%;  Score 478.5;  DB 1;  Length 761;
  Best Local Similarity   22.3%;  Pred. No. 3.9e-27;
  Matches  181;  Conservative  133;  Mismatches  293;  Indels  205;  Gaps   31;

Qy        129 YGMYSREEELLRERKRIGTVGIASY------------DYHQGSGTFLFQAGSGIYHVKD 175
              | : :|| ::     : :|    |              || :    : : |   || :::
Db         79 YNIETRESYIILSNSTMKSVNATDYGLSPDRQFVYLESDYSK-LWRYSYTATYYIYDLQN 137

Qy        176 GGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDW------IAFIHSNDIWISNIVTRE 229
              |    | :   |  ::  |              |       :|::: |:|::
Db        138 G---EFVRGYELPRPIQYLC--------------WSPVGSKLAYVYQNNIYLKQRPGDP 179

Qy        230 ERRLTYV--HNELANMEEDARSAGVATFVLQEEF--DRYSGYWWCPKAETTPSGGKILRI 285
              o  ::|  |  : |        |:  :| :||    :|:  || |         || |
Db        180 PFQITYTGRENRIFN--------GIPDWVYEEEMLATKYA-LWWSP-------DGKFL-- 221

Qy        286 LYEENDESEVEIIHVT------SPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGR 339
              | | ::|:: ||  :       |     |  :  ||| |  || |
Db        222 AYVEFNDSDIPIIAYSYYGDGQYP-----RTINIPYPKAGAKNPVVRV----------F 265

Qy        340 IIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPV 399
              |:|      : | |:     | || : :    ::|   :   :  || :
Db        266 IVDTTYPHHVGPMEVPVP--EMIASSDY----YFSWLTWVSSERVCLQWL---------- 309

Qy        400 EDDVMERQRLIESVPDSVTPLIIYEETTDIWININHDIFHV-------------FPQSHE 445
                    : :::|  ||  :  : |    |    :  ||                |   :
Db        310 --------KRVQNV--SVLSICDFREDWHAWECPKNQEHVEESRTGWAGGFFVSTPAFSQ 359

Qy        446 EEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEV 505
              :   :      | |::|:: |                        |  ::  | ||||:||
Db        360 DATSYYKIFSDKDGYKHIHYI---------------------KDTVENAIQITSGKWEA 397

Qy        506 LGRHGSNIQVDEVRRLVYFEGTKDS----------PLEHHLYVVSYVNPGEVTRLTDRG 554
                        :  |  |:||           |   ::| :|  |     :
Db        398 ---------------IYIFRVTQDSLFYSSNEFEGYPGRRNIYRISIGNSPPSKK----- 437

Qy        555 YSHSCCISQH-----CDFFISKYSNQKNPHCVSLY----KLSSPEDDPTCKTKEFWATIL 605
                |:: |    | :: : :| :   : :  |     :|:  |  |  | :  :
Db        438 -----CVTCHLRKERCQYYTASFSYKAKYYALVCYGPGLPISTLHDGRTDQEIQVLEENK 492

Qy        606 DSAGPLPDYTPPEI-FSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNR 664
              :   | :   |::        | | :  :  |       ||||  :: :||||  | | :
Db        493 ELENSLRNIQLPKVEIKKLKDGGLTFWYKMILPPQFDRSKKYPLLIQVYGGPCSQSVKS- 551

Qy        665 FKGVKYFRLNTLASL----GYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720
```

```
                    | :| :  |    | |: ::| ||:  :| ||  |   |:|  |::||:  ::
Db        552 -----VFAVNWITYLASKEGIVIALVDGRGTAFQGDKFLHAVYRKLGVYEVEDQLTAVRK 606

Qy        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780
                  ||| :|: | ||||||:| :||   : :|:  || ||| | :| : |:||:|
Db        607 FI-EMGFIDEERIAIWGWSYGGYVSSLALASGTGLFKCGIAVAPVSSWEYYASIYSERFM 665

Qy        781 GHP--DQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKP 838
              | |  | | : |   :|  :|| | :      ||:||  |:|||| ::: :   || |
Db        666 GLPTKDDNLEHYKNSTVMARAEYFRNVD--YLLIHGTADDNVHFQNSAQIAKALVNAQVD 723

Qy        839 YDLQIYPQERHSIRVPESGEHYELHLLHYLQE 870
              :    |  : | |    |  |   |: |:|::
Db        724 FQAMWYSDQNHGISSGRSQNHLYTHMTHFLKQ 755


RESULT 9
ST13_YEAST
ID   ST13_YEAST     STANDARD;      PRT;   931 AA.
AC   P33894;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Dipeptidyl aminopeptidase A (EC 3.4.14.-) (DPAP A) (YSCIV).
GN   STE13 OR YCI1 OR YOR219C OR YOR50-9.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95066382; PubMed=7975897;
RA   Anna-Arriola S.S., Herskowitz I.;
RT   "Isolation and DNA sequence of the STE13 gene encoding dipeptidyl
RT   aminopeptidase.";
RL   Yeast 10:801-810(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB320;
RA   Flanagan C.A., Thorner J.;
RT   "STE13.";
RL   (In) Gething M.-J., Novick P., Stevens T.H., Rothblatt J. (eds.);
RL   Guidebook to the yeast secretory pathway, pp.1-1, Oxford University
RL   Press, Oxford (1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1679;
RX   MEDLINE=96437977; PubMed=8840505;
RA   Galisson F., Dujon B.;
RT   "Sequence and analysis of a 33 kb fragment from the right arm of
RT   chromosome XV of the yeast Saccharomyces cerevisiae.";
RL   Yeast 12:877-885(1996).
CC   -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC MATURATION OF THE
CC       ALPHA-FACTOR PRECURSOR.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
CC       VACUOLES.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L21944; AAA35119.1; -.
DR   EMBL; U08230; AAA17897.1; -.
DR   EMBL; X92441; CAA63182.1; -.
DR   EMBL; Z75127; CAA99437.1; -.
DR   PIR; A49737; A49737.
DR   MEROPS; S09.005; -.
DR   SGD; S0005745; STE13.
DR   GO; GO:0005802; C:Golgi trans-face; IDA.
DR   GO; GO:0004177; F:aminopeptidase activity; IDA.
DR   GO; GO:0007323; P:peptide pheromone maturation; IDA.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; FALSE_NEG.
KW   Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW   Transmembrane; Glycoprotein; Signal-anchor; Pheromone response.
FT   DOMAIN       1    119       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   120    140       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN     141    931       LUMENAL (POTENTIAL).
FT   ACT_SITE   785    785       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   863    863       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   896    896       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD   377    377       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   814    814       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   931 AA;  107200 MW;  81AF70094093C023 CRC64;

  Query Match             9.9%;  Score 465;  DB 1;  Length 931;
  Best Local Similarity   21.5%;  Pred. No. 5.1e-26;
  Matches  218;  Conservative  150;  Mismatches  380;  Indels  268;  Gaps   41;

Qy          6 ETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMMAKAPH 65
              :: : |  | |     |  |:  | |:  ||  |:||              |
Db         43 QSNENGHTINEIRPTEATIDVTDVPQ-TPFLQEQYS-----------------MRPRRE 83

Qy         66 DFMFVKRNDPDGPHSDRIYYLAMS----------GENRENTLFYSEIPKTINRAAVLMLS 115
               | |   || :  |    ::                | |    | :: |  ::    |:::
Db         84 SFQF---NDIENQHHTHSFFSVNKFNRRWGEWSLPEKRSYVLVFTLIALSVLVLLVILIP 140

Qy        116 WKPL-----------------------------------------LDLFQATL----DYGM 131
              | |                                         :|  |  |     | |:
Db        141 SKLLPTKITRPKTSAGDSSLGKRSFSIENVLNGDFAIPEDTFHFIDPPQRLLGQDSDPGL 200

Qy        132 YSREEELLRERKRIG-TVGIASYDYHQGSGTFLFQ------------------------ 165
              |   :|:      |   :   ::: : |   ||::
Db        201 YFTTKEIDGHTNFIAKQLFDETFEVNLGGNRFLYEGVEFTVSTVQINYKLDKLIFGTNLE 260

Qy        166 -----AGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPK-----------LCPAD 208
                  :  | | :||          |   :|   |  : |                ||
Db        261 SEFRHSSKGFYWIKD----------LNTGNIEPILPPEKSDDNYELGLSKLSYAHFSPA- 309

Qy        209 PDWIAFIHSNDIWISNIVTREERRLTY-VHNELANMEEDARSAGVATFVLQEE-FDRYSG 266
              ::| |:: |::::  : :   :::|     :: | : |         :: :||
Db        310 YNYIYFVYENNLFLQQVNSGVAKKVTEDGSKDIFNAKPD--------WIYEEEVLASDQA 361

Qy        267 YWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLETRRADS-FRYPKTGTANPKV 325
              ||        |   | :   : :    :: :   |: | |   :|:  :||| |  ||:
Db        362 IWW------APDDSKAVFARFNDTSVDDIRLNRYTN-MNEAYLSDTKIKYPKPGFQNPQF 414

Qy        326 TFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTR 385
                  ::: :  ||  |:    |   ||: |        |      : : |  ||:
Db        415 DL---FLVNLQNGIIYSINTG-GQKDSILYNG-------KWISPDTFRFEI-TDRNSKI 461

Qy        386 LQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWI-NIHDIFHVFPQSH 444
              | : :             |:   | |  :| :: : |       : ||    ||  : |:
Db        462 LDVKVY----------DIPSSQML--TVRNTNSNLF------NGWIEKTKDILSIPPKPE 503

Qy        445 EEEIE--FIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGE 502
              : ::  :|       || ||:   ::                          || | :| |
Db        504 LKRMDYGYIDIHADSRGFSHLFYYPTVF--------------------AKEPIQLTKGN 542

Qy        503 WEVLGRH--GSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYV-----------NPGEVT 548
              |||  |   |  : |     ::|   :   :  ||| :|              ||
Db        543 WEVTGNGIVGYEYETD----TIFFTANEIGVMSQHLYSISLTDSTTQNTFQSLQNP---- 594

Qy        549 RLTDRGYSHSCCISQHCDFFISKYSNQKNP---------------HCVSLYKLSSPEDD 592
              :|:   :   :|    : |||       |                |  |: :|:  |
Db        595 --SDKYDFYDFELSSSARYAISKKLGPDTPIKVAGPLTRVLNVAEIHDDSILQLTKDE-- 650
```

```
Qy        593 PTCKTKEFWATILDSAGPLPDYTPPEIFSFESTT---GFTLYGMLYKPHDLQPGKKYPTV 649
Db        651 ---KFKE----------KIKNYDLP-ITSYKTMVLDDGVEINYIEIKPANLNPKKKYPIL 696

Qy        650 LFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQI 709
Db        697 VNIYGGPGSQTFTT--KSSLAFEQAVVSGLDVIVLQIEPRGTGGKGWSFRSWAREKLGYW 754

Qy        710 EIDDQVE-GLQYLASRYDFIDLDRVGIHGWSYGGYLSLMAL-MQRSDIFRVAIAGAPVTL 767
Db        755 EPRDITEVTKKFIQRNSQHIDESKIAIWGWSYGGFTSLKTVELDNGDTFKYAMAVAPVTN 814

Qy        768 WIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSI 827
Db        815 WTLYDSVYTERYMNQPSENHEGYFEVSTIQNFKSFES-LKRLFIVHGTFDDNVHIQNTFR 873

Qy        828 LLSFLVRAG-KPYDLQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
Db        874 LVDQLNLLGLTNYDMHIFPDSDHSIRYHNAQRIVFQKLYYWLRDAFAERFDNTEVL 929


RESULT 10
SEPR_HUMAN
ID   SEPR_HUMAN     STANDARD;      PRT;   760 AA.
AC   Q12884; O00199; Q99998; Q9UID4;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Seprase (EC 3.4.21.-) (Fibroblast activation protein alpha) (Integral
DE   membrane serine protease) (170-kDa melanoma membrane-bound
DE   gelatinase).
GN   FAP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Fibroblast;
RX   MEDLINE=94261645; PubMed=7911242;
RA   Scanlan M.J., Raj B.K.M., Calvo B., Garin-Chesa P., Sanz-Moncasi M.P.,
RA   Healey J.H., Old L.J., Rettig W.J.;
RT   "Molecular cloning of fibroblast activation protein alpha, a member of
RT   the serine protease family selectively expressed in stromal
RT   fibroblasts of epithelial cancers.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:5657-5661(1994).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Melanoma;
RX   MEDLINE=97388251; PubMed=9247085;
RA   Goldstein L.A., Ghersi G., Pineiro-Sanchez M.L., Salamone M., Yeh Y.,
RA   Flessate D., Chen W.-T.;
RT   "Molecular cloning of seprase: a serine integral membrane protease
RT   from human melanoma.";
RL   Biochim. Biophys. Acta 1361:11-19(1997).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND SEQUENCE OF 220-229; 461-472 AND
RP   511-518.
RC   TISSUE=Melanoma;
RX   MEDLINE=97218181; PubMed=9065413;
RA   Pineiro-Sanchez M.L., Goldstein L.A., Dodt J., Howard L., Yeh Y.,
RA   Chen W.-T.;
RT   "Identification of the 170-kDa melanoma membrane-bound gelatinase
RT   (seprase) as a serine integral membrane protease.";
RL   J. Biol. Chem. 272:7595-7601(1997).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Melanoma;
RX   MEDLINE=20112818; PubMed=10644713;
RA   Goldstein L.A., Chen W.-T.;
RT   "Identification of an alternatively spliced seprase mRNA that encodes
RT   a novel intracellular isoform.";
RL   J. Biol. Chem. 275:2554-2559(2000).
RN   [5]
RP   SEQUENCE OF 192-208; 220-240 AND 510-521.
RX   MEDLINE=94327249; PubMed=7519584;
RA   Rettig W.J., Su S.L., Fortunato S.R., Scanlan M.J., Raj B.K.M.,
RA   Garin-Chesa P., Healey J.H., Old L.J.;
RT   "Fibroblast activation protein: purification, epitope mapping and
RT   induction by growth factors.";
RL   Int. J. Cancer 58:385-392(1994).
CC   -!- FUNCTION: May have a role in tissue remodeling during development
CC       and wound healing, and may contribute to invasiveness in malignant
CC       cancers.
CC   -!- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and
CC       type IV collagen, but not native type I or type IV collagen. Does
CC       not cleave laminin, fibronectin, fibrin or casein.
CC   -!- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is
CC       inactive.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Found in cell
CC       surface lamellipodia, invadopodia and on shed vesicles.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1; Synonyms=L;
CC         IsoId=Q12884-1; Sequence=Displayed;
CC         Note=Major isoform;
CC       Name=2; Synonyms=S, Truncated;
CC         IsoId=Q12884-2; Sequence=VSP_005367;
CC   -!- TISSUE SPECIFICITY: Fibroblast-specific.
CC   -!- INDUCTION: In fibroblasts at times and sites of tissue remodeling
CC       during development, tissue repair, and carcinogenesis.
CC   -!- PTM: N-glycosylated.
CC   -!- PTM: The N-terminus may be blocked.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U09278; AAB49652.1; -.
DR   EMBL; U76833; AAC51668.1; -.
DR   EMBL; AF007822; AAF21600.1; -.
DR   MEROPS; S09.007; -.
DR   Genew; HGNC:3590; FAP.
DR   MIM; 600403; -.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR002471; Prol_endopep_ser.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
DR   PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW   Hydrolase; Protease; Serine protease; Transmembrane; Signal-anchor;
KW   Glycoprotein; Alternative splicing.
FT   DOMAIN        1      4       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      5     25       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       26    760       EXTRACELLULAR (POTENTIAL).
FT   ACT_SITE    624    624       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    702    702       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    734    734       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     49     49       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     92     92       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     99     99       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    314    314       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    679    679       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC      1    521       Missing (in isoform 2).
FT                                /FTId=VSP_005367.
FT   CONFLICT    207    207       P -> A (IN REF. 2).
FT   CONFLICT    229    229       K -> T (IN REF. 2).
FT   CONFLICT    354    354       R -> T (IN REF. 2).
```

```
SQ   SEQUENCE   760 AA;  87820 MW;  A0D34B4801BE07EA CRC64;

  Query Match             9.7%;  Score 454;  DB 1;  Length 760;
  Best Local Similarity   24.1%;  Pred. No. 2.4e-25;
  Matches  169;  Conservative  113;  Mismatches  269;  Indels  150;  Gaps   27;

Qy        212 IAPIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEF--DRYSGYWW 269
              :|::: |:|::         ::|:      |  |:    |:  :| :||    :|:  ||
Db        162 LAYVYQNNIYLKQRPGDPPFQITF------NGRENKIFNGIPDWVYEEEMLPTKYA-LWW 214

Qy        270 CPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLET-RRADSFRYPKTGTANPKVTFK 328
              |        || |   | | :: :: :|  :    |   |  :  ||| |  || |
Db        215 SP-------NGKFL--AYAEFNDKDIPVIAYSYYGDEQYPRTINIPYPKAGAKNPVVRI- 264

Qy        329 MSEIMIDAEGRIIDVIDKELIQPFEI-----LFEGVEYIARAGWTPEGKYAWSILLDRSQ 383
                  |||      : | |:     :      | :   |  : :     |
Db        265 ----------FIIDTTYPAYVGPQEVPVPAMIASSDYYFSWLTWVTDERVCLQWL----- 309

Qy        384 TRLQIVLISPELFIPVEDDVMERQRLIESVPDSVT----------PLIIYEETTDIWINI 433
              |:| | :      :   :  :  | : :| |             |:  |:      |:
Db        310 KRVQNVSVLSICDFREDWQTWDCPKTQEHIEESRTGWAGGFFVSRPVFSYDA-----ISY 364

Qy        434 HDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIK 493
              : ||                    | |::|:: |                     |  ::
Db        365 YKIF----------------SDKDGYKHIHYI---------------------KDTVE 385

Qy        494 EEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDR 553
                | ||||:|| :     ||           |  |:||          |     : |::
Db        386 NAIQITSGKWEAI-----NI----------FRVTQDSLFYSSNEFEEYPGRRNIYRISIG 430

Qy        554 GYSHS-CCISQH-----CDFFISKYSNQKNPHCVSLY----KLSSPEDDPTCKTKEFWAT 603
              |  |  |:: |     | :: : :|:    : :  |     :|:  |  | :
Db        431 SYPPSKKCVTCHLRKERCQYYTASFSDYAKYYALVCYGPGIPISTLHDGRTDQE----IK 486

Qy        604 ILDSAGPLPD-----YTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQV 658
              ||:    | :       | |           ||:  :  |      ||||  :: :||||
Db        487 ILEENKELENALKNIQLPKEEIKKLEVDEITLWYKMILPPQFDRSKKYPLLIQVYGGPCS 546

Qy        659 QLVNNRFKGVKYFRLNTLASL----GYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQ 714
              | | :       | :| :: |    | |: ::| ||:  :| |   |   |:|  |::||
Db        547 QSVRS------VFAVNWISYLASKEGMVIALVDGRGTAFQGDKLLYAVYRKLGVYEVEDQ 600

Qy        715 VEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTG 774
              :  ::        |||  |: | |||||||:| :||   : :|:  || |||: | :| :
Db        601 ITAVRKFI-EMGFIDEKRIAIWGWSYGGYVSSLALASGTGLFKCGIAVAPVSSWEYYASV 659

Qy        775 YTERYMGHP--DQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFL 832
              ||||:|| |  | | : |   :|  :|| | :     ||:||  |:|||| ::: :   |
Db        660 YTERFMGLPTKDDNLEHYKNSTVMARAEYFRNVD--YLLIHGTADDNVHFQNSAQIAKAL 717

Qy        833 VRAGKPYDLQIYPQERHSIRVPESG---EHYELHLLHYLQE 870
              | |   :   | : | :   ||    |   |: |:|::
Db        718 VNAQVDFQAMWYSDQNHGL----SGLSTNHLYTHMTHFLKQ 754


RESULT 11
DPP6_BOVIN
ID   DPP6_BOVIN     STANDARD;      PRT;   863 AA.
AC   P42659;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
DE   (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
DE   (Dipeptidyl aminopeptidase-related protein) (DPPX).
GN   DPP6.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S), AND PARTIAL SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=92108018; PubMed=1729689;
RA   Wada K., Yokotani N., Hunter C., Doi K., Wenthold R.J., Shimasaki S.;
RT   "Differential expression of two distinct forms of mRNA encoding
RT   members of a dipeptidyl aminopeptidase family.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
CC       FUNCTION. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
CC       ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
CC       FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=DPPX-L;
CC         IsoId=P42659-1; Sequence=Displayed;
CC       Name=DPPX-S;
CC         IsoId=P42659-2; Sequence=VSP_005364;
CC   -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN. DPPX-L
CC       IS EXPRESSED EXCLUSIVELY IN THE BRAIN WHEREAS DPPX-S IS FOUND IN
CC       BRAIN, KIDNEY, OVARY AND TESTIS.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M76428; AAC41622.1; -.
DR   EMBL; M76429; AAC41623.1; -.
DR   PIR; A41793; A41793.
DR   MEROPS; S09.973; -.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
KW   Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
FT   DOMAIN        1     93       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     94    114       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN      115    863       EXTRACELLULAR (POTENTIAL).
FT   CARBOHYD    171    171       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    402    402       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    469    469       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    533    533       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    564    564       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    811    811       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC      1     79       MASLYQRFTGKINTSRSFPAPPEASRLLGGQGPEEDGAGPK
FT                                PLGAQAPAAAPRERGGGGGAGGRPRFQYQARSDCDDED ->
FT                                MTTAKEPNASGKSVQQQEQ (in isoform DPPX-S).
FT                                /FTId=VSP_005364.
SQ   SEQUENCE   863 AA;  96556 MW;  23DBA792B841A39D CRC64;

  Query Match             9.4%;  Score 440;  DB 1;  Length 863;
  Best Local Similarity   24.4%;  Pred. No. 3.1e-24;
  Matches  176;  Conservative  114;  Mismatches  317;  Indels  114;  Gaps   29;

Qy        171 YHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREE 230
              |:|    | |  |    | :          ||        : ||  |:|:    | ::
Db        216 YYVLSKIPHGDPQSLDPPEVSNAKLQYAGWGPK-----GQQLIFIFENNIYYCAHVGKQA 270

Qy        231 RRLTYVHNELANMEEDARSAGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEE 289
              |:      ::  :|     |:: :: :||    :  :|| |           |: |
Db        271 IRV------VSTGKEGVIYNGLSDWLYEEEILKTHIAHWWSPDG---------TRLAYAT 315

Qy        290 NDESEVEIIHV---TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDK 346
              ::| | :: :   |  :  |  |  : ||| |  || ::      :|   |   |:
Db        316 INDSRVPVMELPTYTGSVYPT--AKPYHYPKAGCENPSISLH----VIGLNGPTHDL--- 366

Qy        347 ELIQPFEILFEGVEYIARAGWTPEGKYA--WSILLDRSQTRLQIVLISPELFIPVEDDVM 404
```

```
Db        367 EMTPPDDPRMREY-YITMVKWATSTKVAVNW---LSRAQN------VS------------ 404

Qy        405 ERQRLIESVPDSVTPLII--YEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRH 462
Db        405 -----ILTLCDATTGVCTKKHEDESEAWL------------HRQNEEPVFS---KDG-RK 443

Qy        463 LYKITSILK--ESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRR 520
Db        444 FFFVRAIPQGGQGKFYHITVSSSQPNSSNDNIQ---SITSGDWDVT----KILSYDEKRS 496

Qy        521 LVYFEGTKDSPLEHHLYVVSYVNPGEVTR------LTDRGYSHSCCISQHCDFFISKYSN 574
Db        497 QIYFLSTEDLPRRRQLYSASTV--GSFNRQCLSCDLVDNCTYFSASFSPGADFFLLKCEG 554

Qy        575 QKNPHCVSLYKLSSPEDDPTCKTKE-FWATILDSAGPLPDYTPPEIFSFESTTGFTLYGM 633
Db        555 PGVP-TVSVHNTTDKKKMFDLETNEHVQKAISDRQMPKVEYRKIE------TDDYNLPIQ 607

Qy        634 LYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCH 693
Db        608 ILKPATFTDTAHYPLLLVVDGTPGSQSVAEKF--AVTWETVMVSSHGAVVVKCDGRGSGF 665

Qy        694 RGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRS 753
Db        666 QGTRLLHEVRRRLGSLEEKDQMEAVRVML-KEPYIDKTRVAVFGKDYGGYLSTYLLPAKG 724

Qy        754 D----IFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRL 809
Db        725 DGQAPVFSCGSALSPITDFKLYASAFSERYLGLHGLDNRAYEMAKVAHRVSAL--EGQQF 782

Qy        810 LLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQ 869
Db        783 LVIHATADEKIHFQHTAELITQLIKGKANYSLQIYPDESHYFSSAALQQHLHRSILGFFV 842

Qy        870 E 870
Db        843 E 843
```

```
RESULT 12
DPP6_MOUSE
ID   DPP6_MOUSE     STANDARD;      PRT;   804 AA.
AC   Q9Z218; Q9QWW2; Q9Z219;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
DE   (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
DE   (Dipeptidyl aminopeptidase-related protein) (DPPX).
GN   DPP6 OR DPP-6.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Embryo;
RX   MEDLINE=99030650; PubMed=9811881;
RA   Hough R.B., Lengeling A., Bedian V., Lo C., Bucan M.;
RT   "Rump white inversion in the mouse disrupts dipeptidyl aminopeptidase-
RT   like protein 6 and causes dysregulation of kit expression.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:13800-13805(1998).
CC   -!- FUNCTION: May be involved in the physiological processes of brain
CC       function. Has no dipeptidyl aminopeptidase activity. The lack of
CC       activity may be due to the substitution of an aspartate residue
CC       for the serine residue in the proposed catalytic triad (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF092507; AAC97366.1; -.
DR   EMBL; AF092506; AAC97365.1; -.
DR   EMBL; AF092505; AAC98381.1; -.
DR   MEROPS; S09.973; -.
DR   MGD; MGI:94921; Dpp6.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
KW   Transmembrane; Glycoprotein; Signal-anchor.
FT   DOMAIN        1     34       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     35     55       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       56    804       EXTRACELLULAR (POTENTIAL).
FT   CARBOHYD    112    112       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    113    113       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    258    258       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    343    343       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    410    410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    474    474       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    505    505       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    752    752       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    440    455       MISSING (IN REF. 1; AAC98381).
FT   CONFLICT    638    638       Q -> P (IN REF. 1; AAC97365).
SQ   SEQUENCE   804 AA;  91260 MW;  09CFCE7AAD8A7168 CRC64;
```

```
  Query Match             9.3%;  Score 435;  DB 1;  Length 804;
  Best Local Similarity   23.0%;  Pred. No. 6.5e-24;
  Matches  178;  Conservative  127;  Mismatches  322;  Indels  148;  Gaps   30;

Qy        138 LLRERKRIGTVGIASYDYHQGSGTFLF--------QAGSGIYHVKDGGPQGFTQQPLRPN 189
Db        116 VLIEGKKIESLRAIRYEISPDKEYVLFSYNVEPVYQHSHTGYYVLSKIPHGDPQSLDPPE 175

Qy        190 LVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNELANMEEDARS 249
Db        176 VSNAKLQYAGWGPK-----GQQLIFIFENNIYYCAHVGKQAIRV------VSTGKEGVIY 224

Qy        250 AGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV---TSPML 305
Db        225 NGLSDWLYEEEILKSHIAHWWSPDG---------TRLAYATINDSRVPLMELPTYTGSVY 275

Qy        306 ETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARA 365
Db        276 PT--VKPYHYPKAGSENPSISLH----VIGLNGPTHDL---EMMPPDDPRMREY-YITMV 325

Qy        366 GWTPEGKYA--WSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLII- 422
Db        326 KWATSTKVAVTW---LNRAQN------VS----------------ILTLCDATTGVCTK 359

Qy        423 -YEETTDIWINIHDIFHVFPQSHEEEI------EFIFASECKTGFR-HLYKITSILKESK 474
Db        360 KHEDESEAWLH---------RQNEEPVFSKDGRKFFFVRAIPQGGRGKFYHIT--VSSSQ 408

Qy        475 YKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEH 534
Db        409 PNSSNDNIQ------------SITSGDWDVT----KILSYDEKRNKIYFLSTEDLPRRR 451

Qy        535 HLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPT 594
Db        452 HLYSANTVDDFNRQCL-------SCDLVENCTYVSASFSHNMD---FFLLKCEGP-GVPT 500

Qy        595 C---------------KTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHD 639
Db        501 VTVHNTTDKRRMFDLEANEEVQKAINDRQMPKIEYRKIEV------EDYSLPMQILKPAT 554
```

```
Qy        640 LQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFE 699
Db        555 FTDTAHYPLLLVVDGTPGSQSVTERFEVT--WETVLVSSHGAVVVKCDGRGSGFQGTKLL 612

Qy        700 GAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSD----I 755
Db        613 QEVRRRLGFLEEKDQMEAVRTML-KEQYIDKTRVAVFGKDYGGYLSTYILPAKGENQGQT 671

Qy        756 FRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGF 815
Db        672 FTCGSALSPITDFKLYASAFSERYLGLHGLDNRAYEMTKLAHRVSAL--EDQQFLIIHAT 729

Qy        816 LDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQE 870
Db        730 ADEKIHFQHTAELITQLIKGKANYSLQIYPDESHYFHSVALKQHLSRSIIGFFVE 784
```

RESULT 13
DPP6_RAT
ID DPP6_RAT STANDARD; PRT; 859 AA.
AC P46101;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
DE (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
DE (Dipeptidyl aminopeptidase-related protein) (DPPX).
GN DPP6.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX *NCBI_TaxID=10116;*
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S).
RC TISSUE=Brain;
RX MEDLINE=92108018; PubMed=1729689;
RA Wada K., Yokotani N., Hunter C., Doi K., Wenthold R.J., Shimasaki S.;
RT "Differential expression of two distinct forms of mRNA encoding
RT members of a dipeptidyl aminopeptidase family.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
CC -!- FUNCTION: May be involved in the physiological processes of brain
CC function. Has no dipeptidyl aminopeptidase activity. The lack of
CC activity may be due to the substitution of an aspartate residue
CC for the serine residue in the proposed catalytic triad.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=DPPX-L;
CC IsoId=P46101-1; Sequence=Displayed;
CC Name=DPPX-S;
CC IsoId=P46101-2; Sequence=VSP_005366;
CC -!- TISSUE SPECIFICITY: DPPX-S IS EXPRESSED IN BRAIN AND SOME
CC PERIPHERAL TISSUES INCLUDING KIDNEY, OVARY, AND TESTIS; IN
CC CONTRAST DPPX-L IS EXPRESSED ALMOST EXCLUSIVELY IN BRAIN.
CC -!- SIMILARITY: Belongs to peptidase family S9B.
CC ---

CC ---
DR EMBL; M76426; AAC42061.1; -.
DR EMBL; M76427; AAC42062.1; -.
DR MEROPS; S09.973; -.
DR InterPro; IPR002469; DPPIV_N_term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00930; DPPIV_N_term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
KW Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.

FT DOMAIN 1 89 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 90 110 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 111 859 EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 560 560 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 75 MASLYQRFTGKINTSRSFPAPPEASHLLGGQGPEEDAGSKP
FT LGPQAQAVAPRERGGAGGRPRFQYQARSDCDEED -> MTT
FT AKEPSASGKSVQQQDQ (in isoform DPPX-*S*).
FT /FTId=VSP_005366.
SQ SEQUENCE 859 AA; 97301 MW; CE26856D26ED126B CRC64;

Query Match 9.2%; Score 431; DB 1; Length 859;
Best Local Similarity 23.0%; Pred. No. 1.4e-23;
Matches 178; Conservative 129; Mismatches 321; Indels 146; Gaps 30;

```
Qy        138 LLRERKRIGTVGIASYDYHQGSGTFLF--------QAGSGIYHVKDGGPQGFTQQPLRPN 189
Db        171 VLIEGKKIESLRAIRYEISPDKEYALFSYNVEPVYQHSHTGYYVLSKIPHGDPQSLDPPE 230

Qy        190 LVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNELANMEEDARS 249
Db        231 VSNAKLQYAGWGPK-----GQQLIFIFENNIYYCAHVGKQAIRV------VSTGKEGVIY 279

Qy        250 AGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV---TSPML 305
Db        280 NGLSDWLYEEEILKSHIAHWWSPDG---------TRLAYATINDSRVPLMELPTYTGSVY 330

Qy        306 ETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARA 365
Db        331 PT--VKPYHYPKAGSENPSISLH----VIGLNGPTHDL---EMMPPDDPRMREY-YITMV 380

Qy        366 GWTPEGKYA--WSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLII- 422
Db        381 KWATSTKVAVTW---LNRAQN------VS----------ILTLCDATTGVCTK 414

Qy        423 -YEETTDIWINIHDIFHVFPQSHEEEI------EFIFASECKTGFR-HLYKITSILKESK 474
Db        415 KHEDESEAWLH---------RQNEEPVFSKDGRKFFFVRAIPQGGRGKFYHIT--VSSSQ 463

Qy        475 YKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEH 534
Db        464 PNSSNDNIQ------------SITSGDWDVT----EILTYDEKRNKLYFLSTEDLPRRR 506

Qy        535 HLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSP----- 589
Db        507 HLYSANTVDDFNRQCL-------SCDLVENCTYVSASFSHNMD---FFLLKCEGPGVPTV 556

Qy        590 -EDDPTCKTKEF--------WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDL 640
Db        557 TVHNTTDKRRMFDLEANEQVQKAIYDRQMPKIEYRKIEV------EDYSLPMQILKPATF 610

Qy        641 QPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEG 700
Db        611 TDTAHYPLLLVVDGTPGSQSVSERFEVT--WETVLVSSHGAVVVKCDGRGSGFQGTKLLH 668

Qy        701 AFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSD----IF 756
Db        669 EVRRRLGFLEEKDQMEAVRTML-KEQYIDKTRVAVFGKDYGGYLSTYILPAKGENQGQTF 727

Qy        757 RVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFL 816
Db        728 TCGSALSPITDFKLYASAFSERYLGLHGLDNRAYEMTKLAHRVSAL--EDQQFLIIHATA 785

Qy        817 DENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQE 870
Db        786 DEKIHFQHTAELITQLIKGKANYSLQIYPDESHYFHSVALKQHLYRSIIGFFVE 839
```

```
RESULT 14
DPP6_HUMAN
ID   DPP6_HUMAN     STANDARD;      PRT;   865 AA.
AC   P42658;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Dipeptidyl aminopeptidase-like protein 6 (Dipeptidylpeptidase VI)
DE   (Dipeptidylpeptidase 6) (Dipeptidyl peptidase IV like protein)
DE   (Dipeptidyl aminopeptidase-related protein) (DPPX).
GN   DPP6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS DPPX-L AND DPPX-S).
RC   TISSUE=Hippocampus;
RX   MEDLINE=93372805; PubMed=8103397;
RA   Yokotani N., Doi K., Wenthold R.J., Wada K.;
RT   "Non-conservation of a catalytic residue in a dipeptidyl
RT   aminopeptidase IV-related protein encoded by a gene on human
RT   chromosome 7.";
RL   Hum. Mol. Genet. 2:1037-1039(1993).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
CC       FUNCTION. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
CC       ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
CC       FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=DPPX-L;
CC         IsoId=P42658-1; Sequence=Displayed;
CC       Name=DPPX-S;
CC         IsoId=P42658-2; Sequence=VSP_005365;
CC   -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M96859; AAA35760.1; -.
DR   EMBL; M96860; AAA35761.1; -.
DR   PIR; I54331; I54331.
DR   MEROPS; S09.973; -.
DR   Genew; HGNC:3010; DPP6.
DR   MIM; 126141; -.
DR   GO; GO:0008239; F:dipeptidyl-peptidase activity; TAS.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
KW   Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
FT   DOMAIN        1     95       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     96    116       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN      117    865       EXTRACELLULAR (POTENTIAL).
FT   CARBOHYD    173    173       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    319    319       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    404    404       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    471    471       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    535    535       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    566    566       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    813    813       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC      1     81       MASLYQRFTGKINTSRSFPAPPEASHLLGGQGPEEDGGAGA
FT                                KPLGPRAQAAAPRERGGGGGGAGGRPRFQYQGRSDGDEED
FT                                -> MTTAKEPSASGKSVQQQEQE (in isoform
FT                                DPPX-S).
FT                                /FTId=VSP_005365.
SQ   SEQUENCE   865 AA;  97588 MW;  14B1AE0E0024464B CRC64;
```

```
  Query Match             9.1%;  Score 427;  DB 1;  Length 865;
  Best Local Similarity   23.4%;  Pred. No. 2.8e-23;
  Matches  172;  Conservative  121;  Mismatches  299;  Indels  144;  Gaps   30;

Qy        171 YHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREE 230
              |:|     | | |     | :            ||         : ||  |:|:    | ::
Db        218 YYVLSKIPHGDPQSLDPPEVSNAKLQYAGWGPK-----GQQLIFIFENNIYYCAHVGKQA 272

Qy        231 RRLTYVHNELANMEEDARSAGVATFVLQEE-FDRYSGYWWCPKAETTPSGGKILRILYEE 289
               |:       ::  :|     |:: :: :||    :  :|| |                |: |
Db        273 IRV------VSTGKEGVIYNGLSDWLYEEEILKTHIAHWWSPDG---------TRLAYAA 317

Qy        290 NDESEVEIIHV---TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDK 346
               ::| | |: :   |  :  |      : ||| |: || ::      :|   |   |:
Db        318 INDSRVPIMELPTYTGSIYPT--VKPYHYPKAGSENPSISLH----VIGLNGPTHDL--- 368

Qy        347 ELIQPFEILFEGVEYIARAGWTPEGKYA--WSILLDRSQTRLQIVLISPELFIPVEDDVM 404
              |:: | :        ||     |   | |  |   |:|:|       :|
Db        369 EMMPPDDPRMREY-YITMVKWATSTKVAVTW---LNRAQN------VS------------ 406

Qy        405 ERQRLIESVPDSVTPLII--YEETTDIWINIHDIFHVFPQSHEEEI------EFIFASEC 456
                  | :: |: | :    :|: :: |::          : :|| :      :| |
Db        407 -----ILTLCDATTGVCTKKHEDESEAWLH---------RQNEEPVFSKDGRKFFFIRAI 452

Qy        457 KTGFR-HLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQV 515
               | |   | ||  :  |:   |:  :                :||||:|:|       :
Db        453 PQGGRGKFYHIT--VSSSQPNSSNDNIQ------------SITSGDWDVT----KILAY 493

Qy        516 DEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSH---SCCISQHCDFFISKY 572
              ||    :||  |:| |    ||  :           |:  ::    || : ::| :| : :
Db        494 DEKGNKIYFLSTEDLPRRRQLYSAN----------TEGNFNRQCLSCDLVENCTYFSASF 543

Qy        573 SNQKNPHCVSLYKLSSP------EDDPTCKTKEF--------WATILDSAGPLPDYTPPE 618
              |:  :     | |   |        : | | | |          | |   |  :|   |
Db        544 SHSMD---FFLLKCEGPGVPMVTVHNTTDKKKMFDLETNEHVKKAINDRQMPKVEYRDIE 600

Qy        619 IFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLAS 678
              |       : |  : ||       || :| : | |  | |  :|:    :    ::|
Db        601 I------DDYNLPMQILKPATFTDTTHYPLLLVVDGTPGSQSVAEKFE--VSWETVMVSS 652

Qy        679 LGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGW 738
               | |||  | |||  :| |      : ::| :|  ||:| :: :  :  :|| | :  | |
Db        653 HGAVVVKCDGRGSGFQGTKLLHEVRRRLGLLEEKDQMEAVRTML-KEQYIDRTRVAVFGK 711

Qy        739 SYGGYLSLMALMQRSD----IFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGS 794
              ||||||    |  : :      |     | :|:| :  | : ::|||:|    : : | :
Db        712 DYGGYLSTYILPAKGENQGQTFTCGSALSPITDFKLYASAFSERYLGLHGLDNRAYEMTK 771

Qy        795 VAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVP 854
              || :      |  : |::|    || :|| ||: |:: |:|    | |||| | |
Db        772 VAHRVSAL--EEQQFLIIHPTADEKIHFQHTAELITQLIRGKANYSLQIYPDESHYFTSS 829

Qy        855 ESGEHYELHLLLHYLQE 870
               :|    ::::  |
Db        830 SLKQHLYRSIINFFVE 845
```

```
RESULT 15
YDZF_SCHPO
ID   YDZF_SCHPO     STANDARD;      PRT;   853 AA.
AC   Q9P7E9; O13720;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Putative dipeptidyl aminopeptidase C14C4.15c (EC 3.4.14.-).
GN   SPAC14C4.15C OR SPAPJ760.01C.
OS   Schizosaccharomyces pombe (Fission yeast).
```

```
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Lysosome-like
CC       vacuoles.
CC   -!- SIMILARITY: Belongs to peptidase family S9B.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL162631; CAB83084.1; -.
DR   EMBL; Z98596; CAB11208.1; -.
DR   GeneDB_SPombe; SPAC14C4.15c; -.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
KW   Hypothetical protein; Hydrolase; Aminopeptidase; Dipeptidase;
KW   Serine protease; Transmembrane; Glycoprotein; Signal-anchor.
FT   DOMAIN        1     66       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     67     89       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       90    853       LUMENAL (POTENTIAL).
FT   ACT_SITE    719    719       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    795    795       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    828    828       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD     96     96       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    102    102       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    472    472       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    613    613       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   853 AA;  98341 MW;  38450BA50F8304B6 CRC64;

  Query Match             6.2%;  Score 293;  DB 1;  Length 853;
  Best Local Similarity   22.0%;  Pred. No. 1.9e-13;
  Matches 179;  Conservative  110;  Mismatches  308;  Indels  216;  Gaps   38;

Qy        156 YHQGSGTFLFQAGSGIYH----------VKDGGPQGFTQQPLRPNLVETSCPNIRMDPKL 204
              |||     |: :|:           ||   |     |  : | | | |:     |
Db        124 YHQDIEWITSTEGTVLYYDQSTFSLSLFYPDGKEYGSNVDSLITSFVLT-CKNLHR--KR 180

Qy        205 CPADPDWIAFIHSND-IWISN------IVTREERRLTYVHNELA--------------- 241
               :| ::|||  | |  |  :       :| |    |: :: :: :
Db        181 YSSDMEYIAFSCSKDRRWRHSYYEDVYLVERATGRIEHLASDQSKKIVVAEWSPIGHKLV 240

Qy        242 ---------------------NMEEDARSAGVATFVLQEEFDRYS-GYWWCPKAETTPS 278
                                   :  |    |  : :| :||  :  |   || |
Db        241 YGLGSNLFIWESFSEPPVCITDQSDLDGLFNGNSDWVYEEEILQSSKAVWWSPDGNC--- 297

Qy        279 GGKILRILYEENDESEVEIIHV------TSPMLETRRADSFRYPKTGTANPKVTFKMSEI 332
                    : |   |:|:|  :||      | : :  | ::| :  |    ||      ::
Db        298 ------LSYLSIDDSKVP-VHVLPFEQLDSKVEDQNRVNNFFHYST----PKDPIPFVKL 346

Qy        333 MIDA---EGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIV 389
              ::      |  |:|           | :  :   ||    |     | :  :|  :  |:
Db        347 FVNCFTDSGESIEVDSS-----FPLSTQH-RYITDVAWAGNEKLMFVEVLRGNYERV--- 397

Qy        390 LISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHV-------FPQ 442
                     ||     |:  |:  ||:              :::  : :        :
Db        398 ---TSLF-----DLSSRKTTIENTEVSEHPLALTSSLHLKYLSFESLGNLKERYVRQYFL 449

Qy        443 SHEEEIEF---------IFASECKTGF-RHLYKITSILKESKYKRSSGGLPAPSDF----K 489
              |:::  |          |: :     |     ||| | :|     :   | | |  :   |
Db        450 SNKKRIAIYELDNPVPIYLTPVNISFLSDLYLINNTL---YFTAISSGSPFSRVYRLCTK 506

Qy        490 CPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTR 549
                 |  |||  |  |      |   |   :   |:                  :  :|:|: |  |: :
Db        507 SLILSEINIQIGS----GLFGIKVSNDQ-----------------NYLLVNYLGP-EIPK 544

Qy        550 LTDRGYS-HSCCISQHCDFFISKYSNQKNPHCVSLYKL------SSPEDDPTCKTKEFWA 602
                ||  |   :|   |    ||  :    :    || |:      |    ::    ||| :|
Db        545 --QFIYSIHEDKVSTSNDH--SKNNLPSDSSSTSLGKVKLELCNSLETNEELIITKEKFA 600

Qy        603 TILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVN 662
                                     | |      :| :   | |:||||  :|| ||  ||
Db        601 -------------FPSVFFKVIKVKNITAYIQEIRPPNFNPRKRYPTVFHLYGAPQSALVT 648

Qy        663 NRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEG----- 717
               :::      :| |  :  |  :||                    |:  ::| | | |
Db        649 GKYE----MDINELMASVYNFLVI-------------------KVDIVDISD-VSGQHLFS 685

Qy        718 ----------LQYLASRYD--FIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPV 765
                        :: |  |   |  :||   |||| |||:||||:|  ::: :|   :     |||
Db        686 DSHELIIKSWIELLRSYVDTPYIDRHRVGIWGWSFGGYLTL-KILENADFIKTGAVVAPV 744

Qy        766 TLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHT 825
              | | :|| |:| :|   :   |   :   :| |   :  ||:|||  |:|||  :|
Db        745 TDWRYYDAYYSENLLGAYSKQTTAIYDKTAVHYSENF-RKLCGLLVLHGTSDDNVHIENT 803

Qy        826 SILLSFLVRAG----KPYDLQIYPQERHSIRVP 854
                |  :| |       |:  | |  |   |
Db        804 MQLTKAMVEKGVYNYYPF---IVPNANHEFSDP 833

Search completed: October 15, 2003, 17:11:26
Job time : 20.9931 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 15, 2003, 17:02:16 ; Search time 56.4095 Seconds
(without alignments)
4034.822 Million cell updates/sec

Title: US-10-070-464-1
Perfect score: 4700
Sequence: 1 MAAAMETEQLGVEIFETADC.....HLLHYLQENLGSRIAALKVI 882

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4700	100.0	882	4	Q9HBM5	Q9hbm5 homo sapien
2	4696	99.9	882	4	Q8IWG7	Q8iwg7 homo sapien
3	4528.5	96.4	892	11	Q9D4G6	Q9d4g6 mus musculu
4	4397.5	93.6	831	4	Q8NEM5	Q8nem5 homo sapien
5	3333.5	70.9	632	4	Q96JX1	Q96jx1 homo sapien
6	2870	61.1	863	4	Q8WXD8	Q8wxd8 homo sapien
7	2833	60.3	862	11	Q8BWT9	Q8bwt9 mus musculu
8	2823	60.1	862	11	Q8BVG4	Q8bvg4 mus musculu
9	2483	52.8	469	4	Q9NXF4	Q9nxf4 homo sapien
10	2422	51.5	465	4	Q9HBM3	Q9hbm3 homo sapien
11	2068	44.0	628	4	Q8N3F5	Q8n3f5 homo sapien
12	1836.5	39.1	360	4	Q9HBM2	Q9hbm2 homo sapien
13	1645.5	35.0	310	4	Q9HBM4	Q9hbm4 homo sapien
14	1643.5	35.0	508	4	O75273	O75273 homo sapien
15	1610.5	34.3	439	4	Q9BVR3	Q9bvr3 homo sapien
16	1599.5	34.0	1042	5	Q9VC20	Q9vc20 drosophila
17	1599.5	34.0	1102	5	Q9VC19	Q9vc19 drosophila
18	1596	34.0	1053	5	Q8IH07	Q8ih07 drosophila
19	1175	25.0	312	4	Q96NT8	Q96nt8 homo sapien
20	1049	22.3	432	4	O75868	O75868 homo sapien
21	924.5	19.7	927	5	Q965K3	Q965k3 caenorhabdi
22	922.5	19.6	931	5	O44987	O44987 caenorhabdi
23	846	18.0	763	16	Q8EAB7	Q8eab7 shewanella
24	842	17.9	738	16	Q9A6E0	Q9a6e0 caulobacter
25	837.5	17.8	746	10	Q9FNF6	Q9fnf6 arabidopsis
26	754	16.0	741	2	P95782	P95782 xanthomonas
27	751.5	16.0	751	16	Q8P3V8	Q8p3v8 xanthomonas
28	744.5	15.8	757	16	Q8PFD7	Q8pfd7 xanthomonas
29	727	15.5	552	10	Q8GUJ7	Q8guj7 arabidopsis
30	664	14.1	167	4	Q8N2J7	Q8n2j7 homo sapien
31	620	13.2	711	2	Q47900	Q47900 flavobacter
32	582.5	12.4	771	3	O42812	O42812 aspergillus
33	572.5	12.2	723	2	O66223	O66223 porphyromon
34	563.5	12.0	723	2	O31048	O31048 porphyromon
35	553.5	11.8	901	3	Q96VT7	Q96vt7 aspergillus
36	546	11.6	730	2	Q93JY4	Q93jy4 prevotella
37	539.5	11.5	765	3	O14425	O14425 aspergillus
38	537.5	11.4	748	13	P70092	P70092 xenopus lae
39	484.5	10.3	761	11	Q8R492	Q8r492 rattus norv
40	474	10.1	707	16	Q9F348	Q9f348 streptomyce
41	468.5	10.0	814	16	Q8G6Q6	Q8g6q6 bifidobacte
42	463.5	9.9	935	5	Q9VMB4	Q9vmb4 drosophila
43	462	9.8	796	4	Q8N608	Q8n608 homo sapien
44	461	9.8	711	4	Q9P236	Q9p236 homo sapien
45	461	9.8	802	5	Q9VUH1	Q9vuh1 drosophila

ALIGNMENTS

RESULT 1
Q9HBM5
ID Q9HBM5 PRELIMINARY; PRT; 882 AA.
AC Q9HBM5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Dipeptidyl peptidase 8.
GN DPP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
RA Gorrell M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
RT dipeptidyl peptidase (DPP) IV homolog, DPP8.";
RL Eur. J. Biochem. 267:6140-6150(2000).
DR EMBL; AF221634; AAG29766.1; -.
DR MEROPS; S09.018; -.
DR InterPro; IPR002469; DPPIV_N_term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00930; DPPIV_N_term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
SQ SEQUENCE 882 AA; 101421 MW; AD801C302DB4652B CRC64;

Query Match 100.0%; Score 4700; DB 4; Length 882;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 882; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
```

```
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660

Qy        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720

Qy        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780

Qy        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840

Qy        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
              |||||||||||||||||||||||||||||||||||||||||||
Db        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
```

RESULT 2
Q8IWG7
ID Q8IWG7 PRELIMINARY; PRT; 882 AA.
AC Q8IWG7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to dipeptidylpeptidase 8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC040203; AAH40203.1; -.
SQ SEQUENCE 882 AA; 101391 MW; 88C76AF5BCE707F9 CRC64;

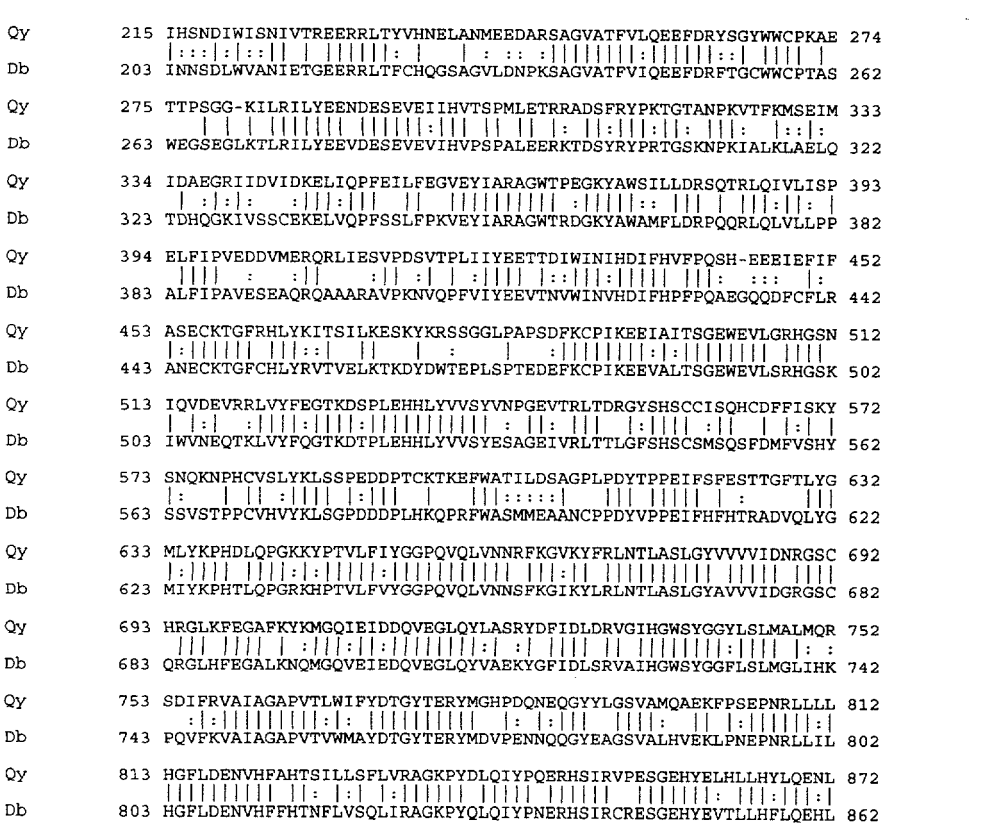
```
  Query Match             99.9%;  Score 4696;  DB 4;  Length 882;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches  881;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHGNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660

Qy        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720

Qy        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780

Qy        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840

Qy        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
              |||||||||||||||||||||||||||||||||||||||||||
Db        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
```

RESULT 3
Q9D4G6
ID Q9D4G6 PRELIMINARY; PRT; 892 AA.
AC Q9D4G6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)

```
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Adult male testis cDNA, RIKEN full-length enriched library,
DE   clone:4932434F09 product:DIPEPTIDYL PEPTIDASE 8 homolog.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RA   Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA   Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA   Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA   Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA   Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA   Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA   Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA   Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA   Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA   Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=21085660; PubMed=11217851;
RA   RIKEN FANTOM Consortium;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=99279253; PubMed=10349636;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20499374; PubMed=11042159;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20530913; PubMed=11076861;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
DR   EMBL; AK016546; BAB30295.2; -.
SQ   SEQUENCE   892 AA;  102284 MW;  FDE0DBEDCA4CA346 CRC64;

  Query Match          96.4%;  Score 4528.5;  DB 11;  Length 892;
  Best Local Similarity   95.7%;  Pred. No. 0;
  Matches  845;  Conservative   18;  Mismatches   19;  Indels    1;  Gaps    1;

Qy          1 MAAAMETEQLGVEIFETADCEE-NIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYM 59
              ||||||||||||||||||:||| | ||||||||||||||||||||||||||||||||||
Db         10 MAAAMETEQLGVEIFETAECEEGNGESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYM 69

Qy         60 MAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPL 119
              ||||||||||||| ||| |||||:|||||||||||||||||||||||||||||||||||
Db         70 MAKAPHDFMFVKRTDPDRPHSDRVYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPL 129

Qy        120 LDLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQ 179
              ||||||||||||||||||||||||||||||:|||| ||||||||||||||:|||||
Db        130 LDLFQATLDYGMYSREEELLRERKRIGTVGIAAYDYHPGSGTFLFQAGSGIYHIKDGGPH 189

Qy        180 GFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNE 239
              |||||||||||||||||||||||||||||||||||||||||||:|||||||:||||||
Db        190 GFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNLVTREERRITYVHNE 249

Qy        240 LANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIH 299
              ||||||| ||||||||||||||||||||||||:|| ||||||||||||||||||||||
Db        250 LANMEEDPRSAGVATFVLQEEFDRYSGYWWCPQAERTPSGGKILRILYEENDESEVEIIH 309

Qy        300 VTSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGV 359
              ||||||||||||||||||||||||||||||||::|| | |||||||||:||||||||||
Db        310 VTSPMLETRRADSFRYPKTGTANPKVTFKMSEIVVDAAGGIIDVIDKELVQPFEILFEGV 369

Qy        360 EYIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTP 419
              |||||||||||:|||||||||| |||||||||||||| |:||||||||||||||
Db        370 EYIARAGWTPEGKHAWSILLDRSQTHLQIVLISPELFIPVEDDAMDRQRLIESVPDSVTP 429

Qy        420 LIIYEETTDIWININHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSS 479
              ||||| |||||||||||||||||:||:||||||||||||||||||||||||||||||||
Db        430 LIIYEETTDIWININHDIFHVFPQTHEDEIEFIFASECKTGFRHLYKITSILKESKYKRSS 489

Qy        480 GGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVV 539
              |||||||||||||||| |||||||||||||| ||| |:|||||||||||||||||||
Db        490 GGLPAPSDFKCPIKEEITITSGEWEVLGRHGSNIWVDEARKLVYFEGTKDSPLEHHLYVT 549

Qy        540 SYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKE 599
              || ||||| |||||||||||:|:|||||||||||||||||||||||||||  ||||
Db        550 SYANPGEVVRLTDRGYSHSCCLSRHCDFFISKYSNQKNPHCVSLYKLSSPEDDPVHKTKE 609

Qy        600 FWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQ 659
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        610 FWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQ 669

Qy        660 LVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQ 719
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        670 LVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQ 729

Qy        720 YLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERY 779
              ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        730 YLASQYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERY 789

Qy        780 MGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPY 839
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        790 MGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPY 849

Qy        840 DLQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
              |||||||||||||||||||||||||||||||||||||||||||
Db        850 DLQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 892


RESULT 4
Q8NEM5
ID   Q8NEM5       PRELIMINARY;      PRT;   831 AA.
AC   Q8NEM5;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Similar to dipeptidylpeptidase 8.
OS   Homo sapiens (Human).
```

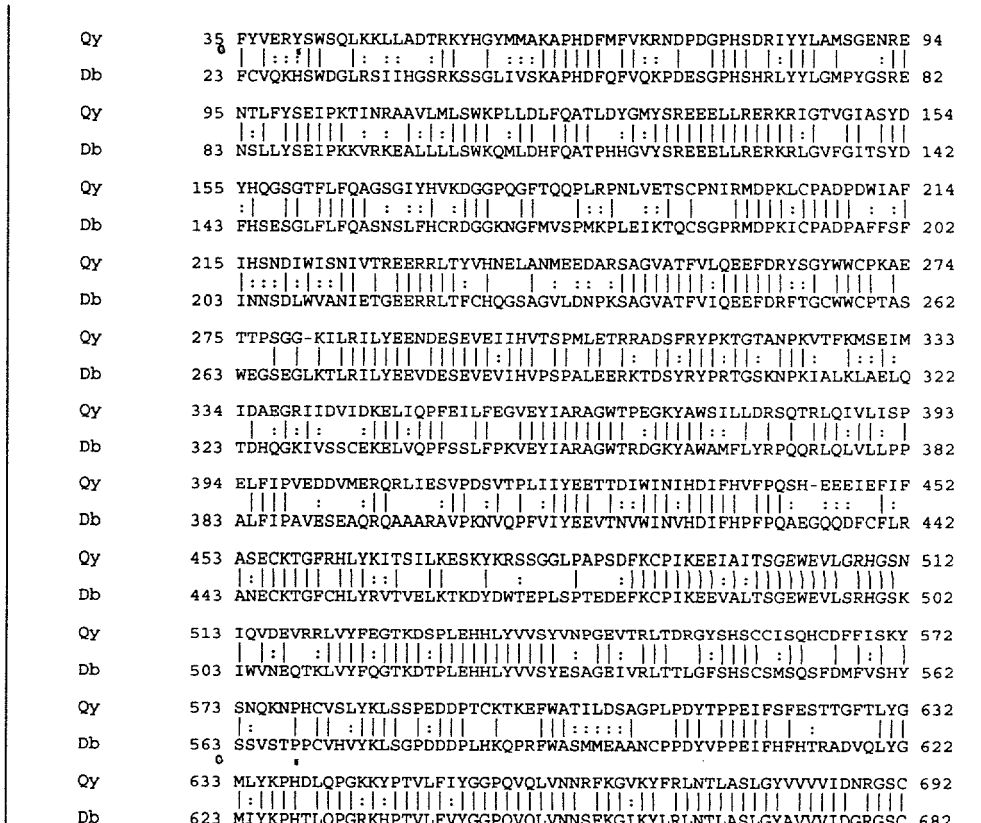
```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Strausberg R.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC030688; AAH30688.1; -.
SQ   SEQUENCE   831 AA;  95527 MW;  0B2A13A2FE70CBE2 CRC64;

  Query Match           93.6%;  Score 4397.5;  DB 4;  Length 831;
  Best Local Similarity  94.2%;  Pred. No. 0;
  Matches  831;  Conservative    0;  Mismatches    0;  Indels   51;  Gaps    1;

Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSG 480

Qy        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVS 540

Qy        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 YVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEF 600

Qy        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 WATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQL 660

Qy        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQY 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 VNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKM-------------- 706

Qy        721 LASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYM 780
                                                ||||||||||||||||||||||||||||
Db        707 ------------------------------------VAIAGAPVTLWIFYDTGYTERYM 729

Qy        781 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        730 GHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD 789

Qy        841 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
              |||||||||||||||||||||||||||||||||||||||||||
Db        790 LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 831


RESULT 5
Q96JX1
ID   Q96JX1      PRELIMINARY;      PRT;   632 AA.
AC   Q96JX1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ14920.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK027826; BAB55395.1; -.
DR   MEROPS; S09.018; -.
KW   Hypothetical protein.
SQ   SEQUENCE   632 AA;  72639 MW;  9BDF598B06985AA4 CRC64;

  Query Match           70.9%;  Score 3333.5;  DB 4;  Length 632;
  Best Local Similarity  92.4%;  Pred. No. 1.9e-250;
  Matches  631;  Conservative    0;  Mismatches    1;  Indels   51;  Gaps    1;

Qy        200 MDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQE 259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQE 60

Qy        260 EFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTG 319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTG 120

Qy        320 TANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILL 379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILL 180

Qy        380 DRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHV 439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHV 240

Qy        440 FPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAIT 499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAIT 300

Qy        500 SGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSC 559
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        301 SGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRSYSHSC 360

Qy        560 CISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEI 619
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEI 420

Qy        620 FSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASL 679
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 FSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASL 480

Qy        680 GYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWS 739
              |||||||||||||||||||||||||||
Db        481 GYVVVVIDNRGSCHRGLKFEGAFKYKM-------------------------------- 507

Qy        740 YGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQA 799
```

```
                                  ||||||||||||||||||||||||||||||||||||||||
Db        508 ------------------VAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQA 549

Qy        800 EKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEH 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        550 EKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEH 609

Qy        860 YELHLLHYLQENLGSRIAALKVI 882
              |||||||||||||||||||||||
Db        610 YELHLLHYLQENLGSRIAALKVI 632


RESULT 6
Q8WXD8
ID   Q8WXD8      PRELIMINARY;      PRT;   863 AA.
AC   Q8WXD8;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Dipeptidyl peptidase 9.
GN   DPP9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Olsen C., Wagtmann N.;
RT   "Identification and characterization of a novel member of the
RT   dipeptidyl peptidase IV-related family.";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RA   Strausberg R.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF452102; AAL47179.1; -.
DR   EMBL; BC037948; AAH37948.1; -.
DR   MEROPS; S09.019; -.
DR   InterPro; IPR002469; DPPIV_N_term.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00930; DPPIV_N_term; 1.
DR   Pfam; PF00326; Peptidase_S9; 1.
SQ   SEQUENCE   863 AA;  98263 MW;  40FE0B78E26CDED5 CRC64;

  Query Match             61.1%;  Score 2870;  DB 4;  Length 863;
  Best Local Similarity   61.5%;  Pred. No. 3.4e-214;
  Matches  517;  Conservative  134;  Mismatches  187;  Indels    2;  Gaps    2;

Qy         35 FYVERYSWSQLKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRE 94
              | |:::||  |: ::  :||| | :: |||||| ||:: |  |||| |:||| |   :||
Db         24 FQVQKHSWDGLRSIIHGSRKYSGLIVNKAPHDFQFVQKTDESGPHSHRLYYLGMPYGSRE 83

Qy         95 NTLFYSEIPKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYD 154
              |:| |||||| : : |:|:|||| :|| ||||  :|:|||||||||||:|  || |||
Db         84 NSLLYSEIPKKVRKEALLLLSWKQMLDHFQATPHHGVYSREEELLRERKRLGVFGITSYD 143

Qy        155 YHQGSGTFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAF 214
              :|  || ||||| : ::| :|||  ||   |::|  ::| |   |||||:||||| : :|
Db        144 FHSESGLFLFQASNSLFHCRDGGKNGFMVSPMKPLEIKTQCSGPRMDPKICPADPAFFSF 203

Qy        215 IHSNDIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAE 274
              |:::|:|::|| | ||||||: |  |:|: :| :||||||||:||||||::|||||| |
Db        204 INNSDLWVANIETGEERRLTFCHQGLSNVLDDPKSAGVATFVIQEEFDRFTGYWWCPTAS 263

Qy        275 TTPSGG-KILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIM 333
                | | |||||||:||||||:||| || || |: ||:|||:||: |||:  |::|
Db        264 WEGSEGLKTLRILYEEVDESEVEVIHVPSPALEERKTDSYRYPRTGSKNPKIALKLAEFQ 323

Qy        334 IDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLISP 393
              |::|:|:   :|||:|||  ||  |||||||||| :|||||:: ||| |  ||:||: |
Db        324 TDSQGKIVSTQEKELVQPFSSLFPKVEYIARAGWTRDGKYAWAMFLDRPQQWLQLVLLPP 383

Qy        394 ELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSH-EEEIEFIF 452
              ||||  ::  :|     :|| :| | ::||| |::|||:|||: ||||  |:|: |:
Db        384 ALFIPSTENEEQRLASARAVPRNVQPYVVYEEVTNVWINVHDIFYPFPQSEGEDELCFLR 443

Qy        453 ASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSN 512
              |:|||||| ||||:|::||   |  |     |  :|||||||||:|||||||| ||||
Db        444 ANECKTGFCHLYKVTAVLKSQGYDWSEPFSPGEDEFKCPIKEEIALTSGEWEVLARHGSK 503

Qy        513 IQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKY 572
              | |:|  :||||:||||:|||||||||||   ||: |||  |:|||| :||: | |:| |
Db        504 IWVNEETKLVYFQGTKDTPLEHHLYVVSYEAAGEIVRLTTPGFSHSCSMSQNFDMFVSHY 563

Qy        573 SNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYG 632
              |:   | || :|||| |:||   |    |||:::::|   ||| ||||| | : :   |||
Db        564 SSVSTPPCVHVYKLSGPDDDPLHKQPRFWASMMEAASCPPDYVPPEIFHFHTRSDVRLYG 623

Qy        633 MLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSC 692
              |:|||| ||||||:|||||:|||||||||| |||:|| |||||||||| ||||| ||||
Db        624 MIYKPHALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNTLASLGYAVVVIDGRGSC 683

Qy        693 HRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQR 752
              |||:|||| | :|||:||:|||||||::| :| |||| || |||||||:|||| |: :
Db        684 QRGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAIHGWSYGGFLSLMGLIHK 743

Qy        753 SDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLL 812
              :|:|||||||||:|: |||||||||| |: |: || ||||:  || |:|||||:|
Db        744 PQVFKVAIAGAPVTVWMAYDTGYTERYMDVPENNQHGYEAGSVALHVEKLPNEPNRLLIL 803

Qy        813 HGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQENL 872
              |||||||||| ||: |:| |:|||||| ||||| |||||| |||||||: |||:||| |
Db        804 HGFLDENVHFFHTNFLVSQLIRAGKPYQLQIYPNERHSIRCPESGEHYEVTLLHFLQEYL 863


RESULT 7
Q8BWT9
ID   Q8BWT9      PRELIMINARY;      PRT;   862 AA.
AC   Q8BWT9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Dipeptidyl peptidase 9 homolog.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Liver;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK050021; BAC34034.1; -.
SQ   SEQUENCE   862 AA;  98001 MW;  B1D566E824A834E8 CRC64;

  Query Match             60.3%;  Score 2833;  DB 11;  Length 862;
  Best Local Similarity   60.8%;  Pred. No. 2.6e-211;
  Matches  511;  Conservative  134;  Mismatches  193;  Indels    2;  Gaps    2;

Qy         35 FYVERYSWSQLKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRE 94
              | |:::||  |: ::  :|| | :::|||||| ||:: |  |||| |:||| |   :||
Db         23 FCVQKHSWDGLRSIIHGSRKSSGLIVSKAPHDFQFVQKPDESGPHSHRLYYLGMPYGSRE 82

Qy         95 NTLFYSEIPKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYD 154
              |:| |||||| : : |:|:|||| :|| ||||  :|:|||||||||||||:|  || |||
Db         83 NSLLYSEIPKKVRKEALLLLSWKQMLDHFQATPHHGVYSREEELLRERKRLGVFGITSYD 142

Qy        155 YHQGSGTFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAF 214
              :|  || ||||| : ::| :|||  ||   |::|  ::| |   |||||:||||| : :|
Db        143 FHSESGLFLFQASNSLFHCRDGGKNGFMVSPMKPLEIKTQCSGPRMDPKICPADPAFFSF 202
```



RESULT 8
Q8BVG4
ID Q8BVG4 PRELIMINARY; PRT; 862 AA.
AC Q8BVG4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Dipeptidyl peptidase 9 homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK078301; BAC37211.1; -.
SQ SEQUENCE 862 AA; 98049 MW; B4FD3476B6F50030 CRC64;

Query Match 60.1%; Score 2823; DB 11; Length 862;
Best Local Similarity 60.7%; Pred. No. 1.5e-210;
Matches 510; Conservative 134; Mismatches 194; Indels 2; Gaps 2;



RESULT 9
Q9NXF4
ID Q9NXF4 PRELIMINARY; PRT; 469 AA.
AC Q9NXF4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ20283.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,

```
RA   Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,
RA   Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK000290; BAA91059.1; -.
DR   MEROPS; S09.018; -.
KW   Hypothetical protein.
SQ   SEQUENCE   469 AA;  54367 MW;  088EED0B1E46C11F CRC64;
```

```
  Query Match           52.8%;  Score 2483;  DB 4;  Length 469;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-184;
  Matches  469;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQLKKLLADTRKYHGYMM 60

Qy         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTINRAAVLMLSWKPLL 120

Qy        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSGTFLFQAGSGIYHVKDGGPQG 180

Qy        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSNDIWISNIVTREERRLTYVHNEL 240

Qy        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHV 300

Qy        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVE 360

Qy        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPL 420

Qy        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSI 469
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 IIYEETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSI 469
```

```
RESULT 10
Q9HBM3
ID   Q9HBM3      PRELIMINARY;      PRT;   465 AA.
AC   Q9HBM3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Dipeptidyl peptidase 8 (Fragment).
GN   DPP8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=20467194; PubMed=11012666;
RA   Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
RA   Gorrell M.D.;
RT   "Cloning, expression and chromosomal localization of a novel human
RT   dipeptidyl peptidase (DPP) IV homolog, DPP8.";
RL   Eur. J. Biochem. 267:6140-6150(2000).
DR   EMBL; AF221636; AAG29768.1; -.
DR   MEROPS; S09.018; -.
FT   NON_TER       1      1
SQ   SEQUENCE   465 AA;  53197 MW;  22939EC0A4FE29CE CRC64;
```

```
  Query Match           51.5%;  Score 2422;  DB 4;  Length 465;
  Best Local Similarity  82.3%;  Pred. No. 9.2e-180;
  Matches  465;  Conservative    0;  Mismatches    0;  Indels  100;  Gaps    1;

Qy        318 TGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSI 377
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSI 60

Qy        378 LLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIF 437
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIF 120

Qy        438 HVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIA 497
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 HVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIA 180

Qy        498 ITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSH 557
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSH 240

Qy        558 SCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPP 617
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPP 300

Qy        618 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLA 677
              |||||||||||||||||||||||||||||||||||||||
Db        301 EIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQ---------------------- 340

Qy        678 SLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHG 737
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ---------------------------------------------------------- 340

Qy        738 WSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAM 797
                                    ||||||||||||||||||||||||||||||||||||||
Db        341 --------------------VAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAM 380

Qy        798 QAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESG 857
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 QAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPESG 440

Qy        858 EHYELHLLHYLQENLGSRIAALKVI 882
              |||||||||||||||||||||||||
Db        441 EHYELHLLHYLQENLGSRIAALKVI 465
```

```
RESULT 11
Q8N3F5
ID   Q8N3F5      PRELIMINARY;      PRT;   628 AA.
AC   Q8N3F5;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   DKFZP762M2413.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RA   Ansorge W., Wirkner U., Mewes H.W., Weil B., Wiemann S.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL834376; CAD39039.1; -.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00326; Peptidase_S9; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   628 AA;  71368 MW;  DC0B6A9440507C3B CRC64;
```

```
  Query Match           44.0%;  Score 2068;  DB 4;  Length 628;
  Best Local Similarity  60.6%;  Pred. No. 5e-152;
```

```
  Matches  380;  Conservative   89;  Mismatches  156;  Indels    2;  Gaps    2;

Qy        220 IWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSG 279
              :|::|| | ||||||: |  |:|: :| :||||||||:||||||::|||||| |     |
Db          1 LWVANIETGEERRLTFCHQGLSNVLDDPKSAGVATFVIQEEFDRFTGYWWCPTASWEGSE 60

Qy        280 G-KILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEG 338
              | | ||||||| ||||||:||| || || |: ||:|||:||: |||:  |::|   |::|
Db         61 GLKTLRILYEEVDESEVEVIHVPSPALEERKTDSYRYPRTGSKNPKIALKLAEFQTDSQG 120

Qy        339 RIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIP 398
              :|:   :|||:|||  ||  |||||||||| :|||||:: ||| |  ||:||: | ||||
Db        121 KIVSTQEKELVQPFSSLFPKVEYIARAGWTRDGKYAWAMFLDRPQQWLQLVLLPPALFIP 180

Qy        399 VEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSHEE-EIEFIFASECK 457
                ::  :|     :|| :| | ::||| |::|||:|||: ||   |      |      :
Db        181 STENEEQRLASARAVPRNVQPYVVYEEVTNVWINVHDIFYPFPNQRERTSSAFSAPMNAR 240

Qy        458 TGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDE 517
                    |    ||   |  |    |   :|||||||||:|||||||| |||| | |:|
Db        241 PASAICTKSPPFLKSQGYDWSEPFSPGEDEFKCPIKEEIALTSGEWEVLARHGSKIWVNE 300

Qy        518 VRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKN 577
              :||||:||||:|||||||||||   ||: |||  |:|||| :||: | |:| ||:
Db        301 ETKLVYFQGTKDTPLEHHLYVVSYEAAGEIVRLTTPGFSHSCSMSQNFDMFVSHYSSVST 360

Qy        578 PHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKP 637
              | || :|||| |:||  |   |||:::::|   ||| ||||| | : :   ||||:|||
Db        361 PPCVHVYKLSGPDDDPLHKQPRFWASMMEAASCPPDYVPPEIFHFHTRSDVRLYGMIYKP 420

Qy        638 HDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLK 697
              | ||||||:|||||:|||||||||||| |||:|| |||||||||| ||||| |||| |||:
Db        421 HALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNTLASLGYAVVVIDGRGSCQRGLR 480

Qy        698 FEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFR 757
              |||| | :|||:||:|||||||::| :| |||| || |||||||||:|||| |: :  :|:
Db        481 FEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFK 540

Qy        758 VAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLD 817
              |||||||||:|: |||||||||||  |: |: ||  ||||:  || |:|||||:||||||
Db        541 VAIAGAPVTVWMAYDTGYTERYMDVPENNQHGYEAGSVALHVEKLPNEPNRLLILHGFLD 600

Qy        818 ENVHFAHTSILLSFLVRAGKPYDLQIY 844
              ||||| ||: |:| |:|||||| || |
Db        601 ENVHFFHTNFLVSQLIRAGKPYQLQEY 627


RESULT 12
Q9HBM2
ID   Q9HBM2      PRELIMINARY;      PRT;   360 AA.
AC   Q9HBM2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Dipeptidyl peptidase 8 (Fragment).
GN   DPP8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20467194; PubMed=11012666;
RA   Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
RA   Gorrell M.D.;
RT   "Cloning, expression and chromosomal localization of a novel human
RT   dipeptidyl peptidase (DPP) IV homolog, DPP8.";
RL   Eur. J. Biochem. 267:6140-6150(2000).
DR   EMBL; AF221637; AAG29769.1; -.
DR   MEROPS; S09.018; -.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00326; Peptidase_S9; 1.
FT   NON_TER       1      1
FT   NON_TER     360    360
SQ   SEQUENCE   360 AA;  41070 MW;  CF81C0BB61423E22 CRC64;

  Query Match             39.1%;  Score 1836.5;  DB 4;  Length 360;
  Best Local Similarity   67.5%;  Pred. No. 2.2e-134;
  Matches  360;  Conservative    0;  Mismatches    0;  Indels  173;  Gaps    1;

Qy        244 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 EEDARSAGVATFVLQEEFDRYSGYWWCPKAETTPSGGKILRILYEENDESEVEIIHVTSP 60

Qy        304 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRIIDVIDKELIQPFEILFEGVEYIA 363
              ||||||||||||||||||||||||||||||||||||
Db         61 MLETRRADSFRYPKTGTANPKVTFKMSEIMIDAEGRII---------------------- 98

Qy        364 RAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPVEDDVMERQRLIESVPDSVTPLIIY 423
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ---------------------------------------------------------- 98

Qy        424 EETTDIWINIHDIFHVFPQSHEEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLP 483
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ---------------------------------------------------------- 98

Qy        484 APSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVN 543
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ------------------------------VDEVRRLVYFEGTKDSPLEHHLYVVSYVN 127

Qy        544 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 603
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        128 PGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWAT 187

Qy        604 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 663
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 ILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNN 247

Qy        664 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 723
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 RFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLAS 307

Qy        724 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 776
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 RYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYT 360


RESULT 13
Q9HBM4
ID   Q9HBM4      PRELIMINARY;      PRT;   310 AA.
AC   Q9HBM4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Dipeptidyl peptidase 8 (Fragment).
GN   DPP8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=20467194; PubMed=11012666;
RA   Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
RA   Gorrell M.D.;
RT   "Cloning, expression and chromosomal localization of a novel human
RT   dipeptidyl peptidase (DPP) IV homolog, DPP8.";
RL   Eur. J. Biochem. 267:6140-6150(2000).
DR   EMBL; AF221635; AAG29767.1; -.
DR   MEROPS; S09.018; -.
DR   InterPro; IPR001375; Peptidase_S9.
DR   Pfam; PF00326; Peptidase_S9; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   310 AA;  35396 MW;  8E87C34026D9C7AC CRC64;
```

```
  Query Match             35.0%;  Score 1645.5;  DB 4;  Length 310;
  Best Local Similarity   86.4%;  Pred. No. 1.2e-119;
  Matches  310;  Conservative    0;  Mismatches    0;  Indels   49;  Gaps    1;

Qy        524 FEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSL 583
Db          1 FEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNPHCVSL 60

Qy        584 YKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPG 643
Db         61 YKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPG 120

Qy        644 KKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKFEGAFK 703
Db        121 KKYPTVLFIYGGPQ---------------------------------------------- 134

Qy        704 YKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGA 763
Db        135 ---GQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGA 191

Qy        764 PVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFA 823
Db        192 PVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFA 251

Qy        824 HTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 882
Db        252 HTSILLSFLVRAGKPYDLQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI 310
```

RESULT 14
O75273

```
ID   O75273      PRELIMINARY;      PRT;   508 AA.
AC   O75273;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   R26984_1 (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA   Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA   Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA   Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA   Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA   Kronmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA   Kobayashi A., Olsen A.S., Carrano A.V.;
RT   "Sequence analysis of a 2.5 Mb region in 19p13.3.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005594; AAC33801.1; -.
DR   MEROPS; S09.019; -.
DR   InterPro; IPR000379; Ser_estrs_site.
FT   NON_TER       1      1
SQ   SEQUENCE   508 AA;  57750 MW;  2F4D6645BE2D2C89 CRC64;

  Query Match             35.0%;  Score 1643.5;  DB 4;  Length 508;
  Best Local Similarity   57.6%;  Pred. No. 3.8e-119;
  Matches  313;  Conservative   77;  Mismatches  108;  Indels   45;  Gaps    6;

Qy        340 IIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLISPELFIPV 399
Db          1 IVSTQEKELVQPFSSLFPKVEYIARAG-------AWAMFLDRPQQWLQLVLLPPALFIPS 53

Qy        400 EDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSH-EEEIEFIFASECKT 458
Db         54 TENEEQRLASARAVPRNVQPYVVYEEVTNVWINVHDIFYPFPQSEGEDELCFLRANECKT 113

Qy        459 GFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGRHGSNIQVDEV 518
Db        114 GFCHLYKVTAVLKSQGYDWSEPFSPG--------EGEQSLTNAIW-----------VNEE 154

Qy        519 RRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCDFFISKYSNQKNP 578
Db        155 TKLVYFQGTKDTPLEHHLYVVSYEAAGEIVRLTTPGFSHSCSMSQNFDMFVSHYSSVSTP 214

Qy        579 HCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPH 638
Db        215 PCVHVYKLSGPDDDPLHKQPRFWASMMEAA---------KIFHFHTRSDVRLYGMIYKPH 265

Qy        639 DLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYVVVVIDNRGSCHRGLKF 698
Db        266 ALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRLNTLASLGYAVVVIDGRGSCQRGLRF 325

Qy        699 EGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSYGGYLSLMALMQRSDIFRV 758
Db        326 EGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKV 385

Qy        759 AIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDE 818
Db        386 AIAGAPVTVWMAYDTGYTERYMDVPENNQHGYEAGSVALHVEKLPNEPNRLLILHGFLDE 445

Qy        819 NVHFFAHTSILLSFLVRAGKPYDL---------QIYPQERHSIRVPESGEHYELHLLHYLQ 869
Db        446 NVHFFHTNFLVSQLIRAGKPYQLQVALPPVSPQIYPNERHSIRCPESGEHYEVTLLHFLQ 505

Qy        870 ENL 872
Db        506 EYL 508
```

RESULT 15
Q9BVR3

```
ID   Q9BVR3      PRELIMINARY;      PRT;   439 AA.
AC   Q9BVR3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Strausberg R.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC000970; AAH00970.1; -.
DR   MEROPS; S09.019; -.
DR   Genew; HGNC:18648; DPP9.
DR   InterPro; IPR001375; Peptidase_S9.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   Pfam; PF00326; Peptidase_S9; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   439 AA;  49926 MW;  A18BBA9E12092BAF CRC64;

  Query Match             34.3%;  Score 1610.5;  DB 4;  Length 439;
  Best Local Similarity   67.0%;  Pred. No. 1.1e-116;
  Matches  294;  Conservative   57;  Mismatches   87;  Indels    1;  Gaps    1;

Qy        435 DIFHVFPQSH-EEEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIK 493
Db          1 DIFYPFPQSEGEDELCFLRANECKTGFCHLYKVTAVLKSQGYDWSEPFSPGEDEFKCPIK 60

Qy        494 EEIAITSGEWEVLGRHGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDR 553
Db         61 EEIALTSGEWEVLARHGSKIWVNEETKLVYFQGTKDTPLEHHLYVVSYEAAGEIVRLTTP 120

Qy        554 GYSHSCCISQHCDFFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPD 613
```

```
Db        121 GFSHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQPRFWASMMEAASCPPD 180

Qy        614 YTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRL 673
              | ||||| | : :   ||||:|||| |||||:|||||:||||||||||| |||:|| ||
Db        181 YVPPEIFHFHTRSDVRLYGMIYKPHALQPGKKHPTVLFVYGGPQVQLVNNSFKGIKYLRL 240

Qy        674 NTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRV 733
              |||||||| ||||| |||| |||:|||| | :|||:||:|||||||::| :| |||| ||
Db        241 NTLASLGYAVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFIDLSRV 300

Qy        734 GIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLG 793
              |||||||:|||| |: :  :|:||||||||||:|: |||||||||  |: |: ||  |
Db        301 AIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVWMAYDTGYTERYMDVPENNQHGYEAG 360

Qy        794 SVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRV 853
              |||:  || |:||||||:|||||||||| ||: |:| |:|||||| ||||| ||||||
Db        361 SVALHVEKLPNEPNRLLILHGFLDENVHFFHTNFLVSQLIRAGKPYQLQIYPNERHSIRC 420

Qy        854 PESGEHYELHLLHYLQENL 872
              |||||||: |||:||| |
Db        421 PESGEHYEVTLLHFLQEYL 439
```

Search completed: October 15, 2003, 17:13:41
Job time : 61.4095 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:15:46 ; Search time 9629.41 Seconds
(without alignments)
3747.092 Million cell updates/sec

Title: US-10-070-464-1
Perfect score: 4700
Sequence: 1 MAAAMETEQLGVEIFETADC.....HLLHYLQENLGSRIAALKVI 882

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10070464/runat_15102003_113553_24810/app_query.fasta_1.2652
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10070464_@CGN_1_1_13561_@runat_15102003_113553_24810 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4700	100.0	2649	6	AX354795	AX354795 Sequence
2	4700	100.0	2656	9	AY172659	AY172659 Homo sapi
3	4700	100.0	2671	6	AX608725	AX608725 Sequence
4	4700	100.0	3106	6	AX342633	AX342633 Sequence
5	4700	100.0	3127	9	AF221634	AF221634 Homo sapi
6	4700	100.0	3143	6	AX354793	AX354793 Sequence
7	4696	99.9	4535	9	BC040203	BC040203 Homo sapi
8	4680	99.6	4829	6	AX608735	AX608735 Sequence
9	4536.5	96.5	4799	10	BC043124	BC043124 Mus muscu
10	4397.5	93.6	3130	9	BC030688	BC030688 Homo sapi
11	4385.5	93.3	4685	6	AX608745	AX608745 Sequence
12	4385	93.3	4676	6	AX608743	AX608743 Sequence
13	4118	87.6	2842	6	AX405770	AX405770 Sequence
14	4092.5	87.1	4523	6	AX608731	AX608731 Sequence
15	4091	87.0	2778	9	AK000290	AK000290 Homo sapi
16	3970.5	84.5	2510	6	AX338497	AX338497 Sequence
17	3771	80.2	2668	6	AX405771	AX405771 Sequence
18	3661.5	77.9	4309	6	AX608737	AX608737 Sequence
19	3364.5	71.6	2161	6	BD157001	BD157001 Primer fo
20	3364.5	71.6	2161	9	AK027826	AK027826 Homo sapi
21	2870	61.1	2602	9	AY172660	AY172660 Homo sapi
22	2870	61.1	2617	6	AX608727	AX608727 Sequence
23	2870	61.1	3716	6	AX480934	AX480934 Sequence
24	2870	61.1	4219	6	AX608751	AX608751 Sequence
25	2870	61.1	4232	9	BC037948	BC037948 Homo sapi
26	2870	61.1	4295	9	AF452102	AF452102 Homo sapi
27	2870	61.1	4302	6	AX608747	AX608747 Sequence
28	2867	61.0	2676	6	AX524928	AX524928 Sequence
29	2863	60.9	3000	9	AF542510	AF542510 Homo sapi
30	2820.5	60.0	4180	6	AX608759	AX608759 Sequence
31	2820.5	60.0	4263	6	AX608757	AX608757 Sequence
32	2649	56.4	4076	6	AX608755	AX608755 Sequence
33	2649	56.4	4159	6	AX608753	AX608753 Sequence
34	2599.5	55.3	4037	6	AX608763	AX608763 Sequence
35	2599.5	55.3	4120	6	AX608761	AX608761 Sequence
36	2422	51.5	1669	9	AF221636	AF221636 Homo sapi
37	2139.5	45.5	3243	9	HSM805448	AL834376 Homo sapi
38	1836.5	39.1	1083	9	AF221637	AF221637 Homo sapi
39	1730.5	36.8	2261	6	AX713363	AX713363 Sequence
40	1730.5	36.8	2261	9	AK054656	AK054656 Homo sapi
41	1645.5	35.0	1197	9	AF221635	AF221635 Homo sapi
42	1610.5	34.3	2546	9	BC000970	BC000970 Homo sapi
43	1596	34.0	3740	3	BT001499	BT001499 Drosophil
44	1391	29.6	2411	6	AX608749	AX608749 Sequence
c 45	1360	28.9	215711	3	AC008205	AC008205 Drosophil

ALIGNMENTS

RESULT 1

```
AX354795
LOCUS       AX354795               2649 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 3 from Patent WO0179473.
ACCESSION   AX354795
VERSION     AX354795.1  GI:18619528
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Meyers,R.A. and Williamson,M.
  TITLE     21953, a human prolyl oligopeptidase family member and uses thereof
  JOURNAL   Patent: WO 0179473-A 3 25-OCT-2001;
            Millennium Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .2649
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      803 a    514 c    586 g    746 t
ORIGIN

Alignment Scores:
Pred. No.:                0          Length:          2649
Score:                    4700.00    Matches:         882
Percent Similarity:       100.00%    Conservative:    0
Best Local Similarity:    100.00%    Mismatches:      0
Query Match:              100.00%    Indels:          0
DB:                       6          Gaps:            0

US-10-070-464-1 (1-882) x AX354795 (1-2649)

Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys  20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT  60

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr  40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 120

Qy         41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 180

Qy         61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer  80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 240

Qy         81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 300

Qy        101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 360

Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 420

Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 480

Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 540

Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 600

Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GATCCAAAATTATGCCCTGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 660

Qy        221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 720

Qy        241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 780

Qy        261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 840

Qy        281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 900

Qy        301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 960

Qy        321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1020

Qy        341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1080

Qy        361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1140

Qy        381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CGCTCCCAGACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1200

Qy        401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1260

Qy        421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1320

Qy        441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 1380

Qy        461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1440

Qy        481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1500

Qy        501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1560

Qy        521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1620

Qy        541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 1680

Qy        561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
```

```
Db       1681 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 1740

Qy        581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600

Db       1741 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 1800

Qy        601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620

Db       1801 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 1860

Qy        621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640

Db       1861 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 1920

Qy        641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660

Db       1921 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTG 1980

Qy        661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680

Db       1981 GTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGT 2040

Qy        681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700

Db       2041 TATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGC 2100

Qy        701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720

Db       2101 GCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATAT 2160

Qy        721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740

Db       2161 CTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTAT 2220

Qy        741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760

Db       2221 GGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATT 2280

Qy        761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780

Db       2281 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 2340

Qy        781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800

Db       2341 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 2400

Qy        801 LysProProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820

Db       2401 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 2460

Qy        821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840

Db       2461 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 2520

Qy        841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860

Db       2521 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATTAT 2580

Qy        861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880

Db       2581 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 2640

Qy        881 ValIle 882

Db       2641 GTGATA 2646

RESULT 2
AY172659
LOCUS       AY172659                2656 bp    mRNA    linear   PRI 08-JAN-2003
DEFINITION  Homo sapiens dipeptidyl peptidase IV-related protein-1 (DPRP1)
            mRNA, complete cds.
ACCESSION   AY172659
VERSION     AY172659.1  GI:27549549
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2656)
  AUTHORS   Qi,S., Akinsanya,K., Riviere,P. and Junien,J.-L.
  TITLE     Novel Serine Protease Genes Related To DPPIV
  JOURNAL   Patent: US (WO 0231134)-A 18-APR-2002;
REFERENCE   2  (bases 1 to 2656)
  AUTHORS   Qi,S., Akinsanya,K., Riviere,P. and Junien,J.-L.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-2002) Ferring Research Institute, 3550 General
            Atomics Ct., San Diego, CA 92121, USA
FEATURES             Location/Qualifiers
     source          1. .2656
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15q22.1-q22.2"
     gene            1. .2656
                     /gene="DPRP1"
     CDS             8. .2656
                     /gene="DPRP1"
                     /codon_start=1
                     /product="dipeptidyl peptidase IV-related protein-1"
                     /protein_id="AAO17261.1"
                     /db_xref="GI:27549550"
                     /translation="MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQ
                     LKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEI
                     PKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSG
                     TFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSN
                     DIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETT
                     PSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIMI
                     DAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLIS
                     PELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWININHDIFHVFPQSHEEEIEF
                     IFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGR
                     HGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCD
                     FFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFES
                     TTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYV
                     VVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSY
                     GGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQ
                     AEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPES
                     GEHYELHLLHYLQENLGSRIAALKVI"
BASE COUNT      804 a    517 c    588 g    747 t
ORIGIN

Alignment Scores:
Pred. No.:                0              Length:          2656
Score:                    4700.00        Matches:         882
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       9              Gaps:            0

US-10-070-464-1 (1-882) x AY172659 (1-2656)

Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20

Db          8 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 67

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40

Db         68 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 127

Qy         41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60

Db        128 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 187

Qy         61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
```

```
Db        188 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 247
Qy         81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
Db        248 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 307
Qy        101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
Db        308 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 367
Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
Db        368 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 427
Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
Db        428 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 487
Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
Db        488 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 547
Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
Db        548 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 607
Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
Db        608 GATCCAAAATTATGCCCTGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 667
Qy        221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
Db        668 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 727
Qy        241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
Db        728 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 787
Qy        261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
Db        788 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 847
Qy        281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
Db        848 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 907
Qy        301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
Db        908 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 967
Qy        321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
Db        968 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1027
Qy        341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
Db       1028 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1087
Qy        361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
Db       1088 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1147
Qy        381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
Db       1148 CGCTCCCAGACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1207
Qy        401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
Db       1208 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1267
Qy        421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
Db       1268 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1327
```

```
Qy        441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
Db       1328 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 1387
Qy        461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
Db       1388 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1447
Qy        481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
Db       1448 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1507
Qy        501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
Db       1508 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1567
Qy        521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
Db       1568 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1627
Qy        541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
Db       1628 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 1687
Qy        561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
Db       1688 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 1747
Qy        581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
Db       1748 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 1807
Qy        601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
Db       1808 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 1867
Qy        621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db       1868 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 1927
Qy        641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
Db       1928 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTG 1987
Qy        661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db       1988 GTGAATAATCGATTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGT 2047
Qy        681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
Db       2048 TATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGC 2107
Qy        701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
Db       2108 GCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATAT 2167
Qy        721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740
Db       2168 CTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTAT 2227
Qy        741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
Db       2228 GGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATT 2287
Qy        761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
Db       2288 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 2347
Qy        781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
Db       2348 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 2407
```

```
Qy        801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2408 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 2467

Qy        821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2468 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 2527

Qy        841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2528 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATTAT 2587

Qy        861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2588 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 2647

Qy        881 ValIle 882
              ||||||
Db       2648 GTGATA 2653

RESULT 3
AX608725
LOCUS       AX608725               2671 bp    DNA     linear   PAT 17-FEB-2003
DEFINITION  Sequence 2 from Patent WO0231134.
ACCESSION   AX608725
VERSION     AX608725.1  GI:28404298
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Qi,S., Akinsanya,K.O., Riviere,P.J. and Junien,J.L.
  TITLE     Novel serine protease genes related to dppiv
  JOURNAL   Patent: WO 0231134-A 2 18-APR-2002;
            Ferring BV (NL)
FEATURES             Location/Qualifiers
     source          1. .2671
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      805 a    524 c    594 g    748 t
ORIGIN

Alignment Scores:
Pred. No.:              0              Length:         2671
Score:                  4700.00        Matches:        882
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-10-070-464-1 (1-882) x AX608725 (1-2671)

Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          8 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 67

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 127

Qy         41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        128 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 187

Qy         61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 247

Qy         81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 307

Qy        101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 367

Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 427

Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 487

Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        488 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 547

Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        548 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 607

Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        608 GATCCAAAATTATGCCCTGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 667

Qy        221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        668 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 727

Qy        241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        728 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 787

Qy        261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        788 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 847

Qy        281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 907

Qy        301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 967

Qy        321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1027

Qy        341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1087

Qy        361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1147

Qy        381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 CGCTCCCAGACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1207

Qy        401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1267

Qy        421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1268 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1327

Qy        441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1328 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 1387
```

```
Qy       461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
Db      1388 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1447

Qy       481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
Db      1448 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1507

Qy       501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
Db      1508 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1567

Qy       521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
Db      1568 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1627

Qy       541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
Db      1628 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 1687

Qy       561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
Db      1688 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 1747

Qy       581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
Db      1748 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 1807

Qy       601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
Db      1808 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 1867

Qy       621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db      1868 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 1927

Qy       641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
Db      1928 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTG 1987

Qy       661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db      1988 GTGAATAATCGATTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGT 2047

Qy       681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
Db      2048 TATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGC 2107

Qy       701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
Db      2108 GCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATAT 2167

Qy       721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740
Db      2168 CTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTAT 2227

Qy       741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
Db      2228 GGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATT 2287

Qy       761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
Db      2288 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 2347

Qy       781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
Db      2348 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 2407

Qy       801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
Db      2408 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 2467

Qy       821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
Db      2468 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 2527

Qy       841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
Db      2528 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATTAT 2587

Qy       861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
Db      2588 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 2647

Qy       881 ValIle 882
Db      2648 GTGATA 2653
```

```
RESULT 4
AX342633
LOCUS       AX342633               3106 bp    DNA     linear   PAT 12-JAN-2002
DEFINITION  Sequence 30 from Patent WO0198468.
ACCESSION   AX342633
VERSION     AX342633.1  GI:18152030
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Yue,H., Elliott,V.S., Gandhi,A.R., Lal,P., Au-Young,J.,
            Tribouley,C.M., Delegeane,A.M., Baughn,M.R., Nguyen,D.B., Lee,E.A.,
            Hafalia,A., Khan,F.A., Walia,N.K., Yao,M.G., Lu,D.A., Patterson,C.,
            Tang,Y.T., Walsh,R.T., Azimzai,Y., Ramkumar,J., Xu,Y. and Reddy,R.
  JOURNAL   Patent: WO 0198468-A 30 27-DEC-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .3106
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 7160544CB1"
BASE COUNT      928 a    633 c    704 g    841 t
ORIGIN

Alignment Scores:
Pred. No.:              0              Length:         3106
Score:                  4700.00        Matches:        882
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-10-070-464-1 (1-882) x AX342633 (1-3106)

Qy         1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
Db       203 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 262

Qy        21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
Db       263 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 322

Qy        41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
Db       323 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 382

Qy        61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
Db       383 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 442

Qy        81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
Db       443 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 502

Qy       101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
```

```
Db        503 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 562

Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140

Db        563 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 622

Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160

Db        623 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 682

Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180

Db        683 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 742

Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200

Db        743 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 802

Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220

Db        803 GATCCAAAATTATGCCCTGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 862

Qy        221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240

Db        863 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 922

Qy        241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260

Db        923 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 982

Qy        261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280

Db        983 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 1042

Qy        281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300

Db       1043 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 1102

Qy        301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320

Db       1103 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 1162

Qy        321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340

Db       1163 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1222

Qy        341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360

Db       1223 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1282

Qy        361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380

Db       1283 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1342

Qy        381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400

Db       1343 CGCTCCCAGACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1402

Qy        401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420

Db       1403 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1462

Qy        421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440

Db       1463 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1522

Qy        441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460

Db       1523 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 1582

Qy        461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480

Db       1583 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1642

Qy        481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500

Db       1643 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1702

Qy        501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520

Db       1703 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1762

Qy        521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540

Db       1763 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1822

Qy        541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560

Db       1823 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 1882

Qy        561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580

Db       1883 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 1942

Qy        581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600

Db       1943 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 2002

Qy        601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620

Db       2003 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 2062

Qy        621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640

Db       2063 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 2122

Qy        641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660

Db       2123 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTG 2182

Qy        661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680

Db       2183 GTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGT 2242

Qy        681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700

Db       2243 TATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGC 2302

Qy        701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720

Db       2303 GCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATAT 2362

Qy        721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740

Db       2363 CTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTAT 2422

Qy        741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760

Db       2423 GGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATT 2482

Qy        761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780

Db       2483 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 2542

Qy        781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800

Db       2543 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 2602

Qy        801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820

Db       2603 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 2662

Qy        821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840

Db       2663 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 2722
```

```
Qy        841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2723 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATTAT 2782

Qy        861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2783 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 2842

Qy        881 ValIle 882
              ||||||
Db       2843 GTGATA 2848

RESULT 5
AF221634
LOCUS       AF221634               3127 bp    mRNA    linear   PRI 05-NOV-2000
DEFINITION  Homo sapiens dipeptidyl peptidase 8 (DPP8) mRNA, complete cds.
ACCESSION   AF221634
VERSION     AF221634.1  GI:11095187
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3127)
  AUTHORS   Abbott,C.A., Yu,D.M., Woollatt,E., Sutherland,G.R., McCaughan,G.W.
            and Gorrell,M.D.
  TITLE     Cloning, expression and chromosomal localization of a novel human
            dipeptidyl peptidase (DPP) IV homolog, DPP8
  JOURNAL   Eur. J. Biochem. 267 (20), 6140-6150 (2000)
  MEDLINE   20467194
   PUBMED   11012666
REFERENCE   2  (bases 1 to 3127)
  AUTHORS   Abbott,C.A., Yu,D., McCaughan,G.W. and Gorrell,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JAN-2000) A.W. Morrow Gastroenterology and Liver
            Centre, Centenary Institute of Cell Biology and Cancer Medicine,
            Locked Bag No.6, Newtown, Sydney, NSW 2042, Australia
FEATURES             Location/Qualifiers
     source          1. .3127
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15q22"
     gene            1. .3127
                     /gene="DPP8"
     CDS             214. .2862
                     /gene="DPP8"
                     /note="serine protease; member of prolyl oligopetidase S9b
                     family belonging to enzyme clan SC; alternatively spliced
                     forms of this mRNA are found in AF221635, AF221636 and
                     AF221637"
                     /codon_start=1
                     /product="dipeptidyl peptidase 8"
                     /protein_id="AAG29766.1"
                     /db_xref="GI:11095188"
                     /translation="MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQ
                     LKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEI
                     PKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSG
                     TFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSN
                     DIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETT
                     PSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIMI
                     DAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLIS
                     PELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWININHDIFHVFPQSHEEEIEF
                     IFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGR
                     HGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCD
                     FFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFES
                     TTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYV
                     VVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSY
                     GGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQ
                     AEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPES
                     GEHYELHLLHYLQENLGSRIAALKVI"

BASE COUNT      943 a    637 c    706 g    841 t
ORIGIN

Alignment Scores:
Pred. No.:                0              Length:          3127
Score:                    4700.00        Matches:         882
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       9              Gaps:            0

US-10-070-464-1 (1-882) x AF221634 (1-3127)

Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 273

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 333

Qy         41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 393

Qy         61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        394 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 453

Qy         81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 513

Qy        101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 573

Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 633

Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 693

Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        694 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 753

Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        754 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 813

Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        814 GATCCAAAATTATGCCCCGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 873

Qy        221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        874 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 933

Qy        241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        934 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 993

Qy        261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        994 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 1053

Qy        281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1054 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 1113
```

```
Qy        301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 1173

Qy        321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1174 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1233

Qy        341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1234 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1293

Qy        361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1294 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1353

Qy        381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1354 CGCTCCCAGACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1413

Qy        401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1414 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1473

Qy        421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1474 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1533

Qy        441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1534 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 1593

Qy        461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1594 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1653

Qy        481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1654 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1713

Qy        501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1714 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1773

Qy        521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1774 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1833

Qy        541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1834 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 1893

Qy        561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1894 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 1953

Qy        581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1954 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 2013

Qy        601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2014 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 2073

Qy        621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2074 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 2133

Qy        641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2134 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTG 2193

Qy        661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2194 GTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGT 2253

Qy        681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2254 TATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGC 2313

Qy        701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2314 GCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATAT 2373

Qy        721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2374 CTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTAT 2433

Qy        741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2434 GGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATT 2493

Qy        761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2494 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 2553

Qy        781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2554 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 2613

Qy        801 LysProProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2614 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 2673

Qy        821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2674 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 2733

Qy        841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2734 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATTAT 2793

Qy        861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2794 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 2853

Qy        881 ValIle 882
              ||||||
Db       2854 GTGATA 2859
```

```
RESULT 6
AX354793
LOCUS       AX354793               3143 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 1 from Patent WO0179473.
ACCESSION   AX354793
VERSION     AX354793.1  GI:18619526
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Meyers,R.A. and Williamson,M.
  TITLE     21953, a human prolyl oligopeptidase family member and uses thereof
  JOURNAL   Patent: WO 0179473-A 1 25-OCT-2001;
            Millennium Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .3143
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CDS             229. .2877
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD22990.1"
                     /db_xref="GI:18619527"
```

```
                     /translation="MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQ
                     LKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEI
                     PKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSG
                     TFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSN
                     DIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETT
                     PSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIMI
                     DAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLIS
                     PELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWININHDIFHVFPQSHEEEIEF
                     IFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGR
                     HGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCD
                     FFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFES
                     TTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYV
                     VVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSY
                     GGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQ
                     AEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPES
                     GEHYELHLLHYLQENLGSRIAALKVI"
BASE COUNT      943 a    644 c    712 g    844 t
ORIGIN

Alignment Scores:
Pred. No.:              0               Length:         3143
Score:                  4700.00         Matches:        882
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-10-070-464-1 (1-882) x AX354793 (1-3143)

Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        229 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 288

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        289 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 348

Qy         41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        349 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 408

Qy         61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        409 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 468

Qy         81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        469 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 528

Qy        101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        529 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 588

Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        589 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 648

Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        649 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 708

Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        709 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 768

Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        769 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 828

Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        829 GATCCAAAATTATGCCCTGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 888

Qy        221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        889 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 948

Qy        241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        949 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 1008

Qy        261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1009 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 1068

Qy        281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1069 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 1128

Qy        301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1129 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 1188

Qy        321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1189 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1248

Qy        341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1249 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1308

Qy        361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1309 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1368

Qy        381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1369 CGCTCCCAGACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1428

Qy        401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1429 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1488

Qy        421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1489 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1548

Qy        441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1549 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 1608

Qy        461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1609 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1668

Qy        481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1669 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1728

Qy        501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1729 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1788

Qy        521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1789 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1848

Qy        541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1849 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 1908

Qy        561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1909 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 1968

Qy        581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
```

```
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1969 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 2028

Qy       601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2029 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 2088

Qy       621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2089 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 2148

Qy       641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2149 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTG 2208

Qy       661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2209 GTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGT 2268

Qy       681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2269 TATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGC 2328

Qy       701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2329 GCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATAT 2388

Qy       721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2389 CTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTAT 2448

Qy       741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2449 GGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATT 2508

Qy       761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2509 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 2568

Qy       781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2569 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 2628

Qy       801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2629 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 2688

Qy       821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2689 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 2748

Qy       841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2749 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATTAT 2808

Qy       861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2809 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 2868

Qy       881 ValIle 882
             ||||||
Db      2869 GTGATA 2874
```

```
RESULT 7
BC040203
LOCUS       BC040203                4535 bp    mRNA    linear   PRI 03-DEC-2002
DEFINITION  Homo sapiens, similar to dipeptidylpeptidase 8, clone MGC:42570
            IMAGE:4824813, mRNA, complete cds.
ACCESSION   BC040203
VERSION     BC040203.1  GI:26007916
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4535)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 71 Row: k Column: 23.
FEATURES             Location/Qualifiers
     source          1..4535
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:42570 IMAGE:4824813"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
     CDS             1629..4277
                     /codon_start=1
                     /product="similar to dipeptidylpeptidase 8"
                     /protein_id="AAH40203.1"
                     /db_xref="GI:26007917"
                     /db_xref="LocusID:54878"
                     /translation="MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQ
                     LKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEI
                     PKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSG
                     TFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHGN
                     DIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETT
                     PSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIMI
                     DAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLIS
                     PELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSHEEEIEF
                     IFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGR
                     HGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCD
                     FFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFES
                     TTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYV
                     VVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASRYDFIDLDRVGIHGWSY
                     GGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYYLGSVAMQ
                     AEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERHSIRVPES
                     GEHYELHLLHYLQENLGSRIAALKVI"
BASE COUNT     1313 a    918 c    946 g   1358 t
ORIGIN

Alignment Scores:
Pred. No.:              0              Length:         4535
Score:                  4696.00        Matches:        881
Percent Similarity:     99.89%         Conservative:   0
Best Local Similarity:  99.89%         Mismatches:     1
Query Match:            99.91%         Indels:         0
DB:                     9              Gaps:           0

US-10-070-464-1 (1-882) x BC040203 (1-4535)
```

```
Qy       1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
Db    1629 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 1688

Qy      21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
Db    1689 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 1748

Qy      41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
Db    1749 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 1808

Qy      61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
Db    1809 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 1868

Qy      81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
Db    1869 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 1928

Qy     101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
Db    1929 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 1988

Qy     121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
Db    1989 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 2048

Qy     141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
Db    2049 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGG 2108

Qy     161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
Db    2109 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 2168

Qy     181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
Db    2169 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 2228

Qy     201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
Db    2229 GATCCAAAATTATGCCCTGCTGATCCAGACTGGATTGCTTTTATACATGGCAACGATATT 2288

Qy     221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
Db    2289 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 2348

Qy     241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
Db    2349 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 2408

Qy     261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
Db    2409 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 2468

Qy     281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
Db    2469 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 2528

Qy     301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
Db    2529 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 2588

Qy     321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
Db    2589 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 2648

Qy     341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
Db    2649 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 2708

Qy     361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
Db    2709 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 2768

Qy     381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
Db    2769 CGCTCCCAGACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 2828

Qy     401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
Db    2829 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 2888

Qy     421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
Db    2889 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 2948

Qy     441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
Db    2949 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 3008

Qy     461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
Db    3009 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 3068

Qy     481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
Db    3069 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 3128

Qy     501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
Db    3129 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 3188

Qy     521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
Db    3189 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 3248

Qy     541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
Db    3249 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 3308

Qy     561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
Db    3309 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 3368

Qy     581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
Db    3369 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 3428

Qy     601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
Db    3429 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 3488

Qy     621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db    3489 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 3548

Qy     641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
Db    3549 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTG 3608

Qy     661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db    3609 GTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGT 3668

Qy     681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
Db    3669 TATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGC 3728

Qy     701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
Db    3729 GCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATAT 3788

Qy     721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740
```

```
Db       3789 CTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTAT 3848

Qy        741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760

Db       3849 GGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATT 3908

Qy        761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780

Db       3909 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 3968

Qy        781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800

Db       3969 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 4028

Qy        801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820

Db       4029 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 4088

Qy        821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840

Db       4089 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 4148

Qy        841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860

Db       4149 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATTAT 4208

Qy        861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880

Db       4209 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 4268

Qy        881 ValIle 882

Db       4269 GTGATA 4274

RESULT 8
AX608735
LOCUS       AX608735                4829 bp    DNA     linear   PAT 17-FEB-2003
DEFINITION  Sequence 12 from Patent WO0231134.
ACCESSION   AX608735
VERSION     AX608735.1  GI:28404303
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Qi,S., Akinsanya,K.O., Riviere,P.J. and Junien,J.L.
  TITLE     Novel serine protease genes related to dppiv
  JOURNAL   Patent: WO 0231134-A 12 18-APR-2002;
            Ferring BV (NL)
FEATURES             Location/Qualifiers
     source          1. .4829
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT     1466 a    886 c   1017 g   1460 t
ORIGIN

Alignment Scores:
Pred. No.:                0              Length:          4829
Score:                    4680.00        Matches:         882
Percent Similarity:       99.77%         Conservative:    0
Best Local Similarity:    99.77%         Mismatches:      0
Query Match:              99.57%         Indels:          2
DB:                       6              Gaps:            0

US-10-070-464-1 (1-882) x AX608735 (1-4829)

Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20

Db        214 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 273

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40

Db        274 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 333

Qy         41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60

Db        334 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 393

Qy         61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80

Db        394 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 453

Qy         81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100

Db        454 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 513

Qy        101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120

Db        514 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 573

Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140

Db        574 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 633

Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160

Db        634 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 693

Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180

Db        694 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 753

Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200

Db        754 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 813

Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220

Db        814 GATCCAAAATTATGCCCTGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 873

Qy        221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240

Db        874 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 933

Qy        241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260

Db        934 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 993

Qy        261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280

Db        994 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 1053

Qy        281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300

Db       1054 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 1113

Qy        301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320

Db       1114 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 1173

Qy        321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340

Db       1174 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1233

Qy        341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360

Db       1234 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1293

Qy        361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380

Db       1294 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1353

Qy        381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
```

```
Db       1354 CGCTCCCAGACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1413

Qy        401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1414 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1473

Qy        421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1474 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1533

Qy        441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1534 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTGCCTCTGAATGCAAAACAGGTTTC 1593

Qy        461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1594 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1653

Qy        481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1654 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1713

Qy        501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1714 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1773

Qy        521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1774 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1833

Qy        541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1834 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 1893

Qy        561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1894 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 1953

Qy        581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1954 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 2013

Qy        601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2014 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 2073

Qy        621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2074 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 2133

Qy        641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGly--ProGlnValGlnL 660
              |||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db       2134 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCTCCTCAGGTGCAGT 2193

Qy        660 euValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuG 680
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2194 TGGTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAG 2253

Qy        680 lyTyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluG 700
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2254 GTTATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAG 2313

Qy        700 lyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnT 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2314 GCGCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAAT 2373

Qy        720 yrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerT 740
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2374 ATCTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCT 2433

Qy        740 yrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaI 760
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2434 ATGGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTA 2493
```

```
Qy        760 leAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrM 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2494 TTGCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATA 2553

Qy        780 etGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaG 800
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2554 TGGGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAG 2613

Qy        800 luLysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnV 820
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2614 AAAAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATG 2673

Qy        820 alHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrA 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2674 TCCATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATG 2733

Qy        840 spLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisT 860
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2734 ATTTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATT 2793

Qy        860 yrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuL 880
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2794 ATGAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAA 2853

Qy        880 ysValIle 882
              ||||||||
Db       2854 AAGTGATA 2861
```

RESULT 9
BC043124

```
LOCUS       BC043124                4799 bp    mRNA    linear   ROD 22-APR-2003
DEFINITION  Mus musculus dipeptidylpeptidase 8, mRNA (cDNA clone MGC:58057
            IMAGE:6410075), complete cds.
ACCESSION   BC043124
VERSION     BC043124.1  GI:27695449
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4799)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 4799)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
```

```
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            cDNA Library Preparation: M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
            Thomas L. Casavant.
            Web site: http://genome.uiowa.edu
            Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
            Bonaldo,M.F., Akabogu,I., Bair,T., Bair,J., Crouch,K., Davis,A.,
            Fishler,K., Keppel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K.,
            Scheetz,T., Smith,C., Snir,E., Tack,D., Trout,K., Walters,J.,
            Casavant,T., Soares,M.B.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series:  Plate:  Row:  Column: 0.
FEATURES              Location/Qualifiers
     source           1. .4799
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6"
                      /db_xref="taxon:10090"
                      /clone="MGC:58057 IMAGE:6410075"
                      /tissue_type="Brain, enriched mouse brain 12.5dp"
                      /clone_lib="NIH_BMAP_FO0"
                      /lab_host="DH10B"
                      /note="Vector: pYX-ASC"
     gene             1. .4799
                      /gene="Dpp8"
                      /db_xref="LocusID:74388"
                      /db_xref="MGI:1921638"
     CDS              288. .2966
                      /codon_start=1
                      /product="Dpp8 protein"
                      /protein_id="AAH43124.1"
                      /db_xref="GI:27695450"
                      /db_xref="LocusID:74388"
                      /translation="MKIPSGRCNMAAAMETEQLGVEIFETAECEEGNGESQDRPKLEP
                      FYVERYSWSQLKKLLADTRKYHGYMMAKAPHDFMFVKRTDPDGPHSDRVYYLAMSGEN
                      RENTLFYSEIPKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGI
                      AAYDYHPGSGTFLFQAGSGIYHIKDGGPHGFTQQPLRPNLVETSCPNIRMDPKLCPAD
                      PDWIAFIHSNDIWISNLVTREERRITYVHNELANMEEDPRSAGVATFVLQEEFDRYSG
                      YWWCPQAERTPSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPK
                      VTFKMSEIVVDAAGGIIDVIDKELVQPFEILFEGVEYIARAGWTPEGKHAWSILLDRS
                      QTHLQIVLISPELFIPVEDDAMDRQRLIESVPDSVTPLIIYEETTDIWININHDIFHVF
                      PQTHEDEIEFIFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEITI
                      TSGEWEVLGRHGSNIWVDEARKLVYFEGTKDSPLEHHLYVTSYANPGEVVRLTDRGYS
                      HSCCLSRHCDFFISKYSNQKNPHCVSLYKLSSPEDDPVHKTKEFWATILDSAGPLPDY
                      TPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFR
                      LNTLASLGYVVVVIDNRGSCHRGLKFEGAFKYKMGQIEIDDQVEGLQYLASQYDFIDL
                      DRVGIHGWSYGGYLSLMALMQRSDIFRVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQ
                      GYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQ
                      ERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI"
BASE COUNT     1373 a   1015 c   1103 g   1308 t
ORIGIN

Alignment Scores:
Pred. No.:                0              Length:          4799
Score:                    4536.50        Matches:         846
Percent Similarity:       97.85%         Conservative:    18
Best Local Similarity:    95.81%         Mismatches:      18
Query Match:              96.52%         Indels:          1
DB:                       10             Gaps:            1

US-10-070-464-1 (1-882) x BC043124 (1-4799)

Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
Db        315 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTCGAGATATTTGAAACTGCAGAGTGT 374

Qy         21 GluGlu---AsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArg 39
              ||||||   |||   ||||||||||||||||||||||||||||||||||||||||||||||

Db        375 GAGGAGGGAAATGGTGAATCACAGGACCGACCCAAACTGGAGCCTTTTTATGTCGAACGG 434

Qy         40 TyrSerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMet 59
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 TATTCTTGGAGTCAGCTGAAAAAGCTGCTTGCTGATACCAGAAAATACCACGGCTACATG 494

Qy         60 MetAlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHis 79
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db        495 ATGGCTAAGGCGCCACATGACTTTATGTTTGTAAAGAGGACGGATCCAGATGGGCCTCAC 554

Qy         80 SerAspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyr 99
              |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
Db        555 TCAGACAGGGTCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTATTTTAT 614

Qy        100 SerGluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeu 119
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        615 TCTGAAATCCCTAAAACCATCAACAGAGCAGCAGTCTTAATGCTTTCCTGGAAGCCCCTT 674

Qy        120 LeuAspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeu 139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        675 TTGGATCTTTTTCAGGCGACACTAGACTATGGGATGTATTCTCGAGAGGAAGAGCTACTC 734

Qy        140 ArgGluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySer 159
              |||||||||||||||||||||||||||||||||:::|||||||||||||    ||||||
Db        735 AGAGAAAGAAAGCGCATTGGAACCGTGGGGATCGCAGCTTATGATTACCACCCAGGAAGC 794

Qy        160 GlyThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGln 179
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db        795 GGAACATTCCTGTTTCAAGCTGGCAGCGGCATTTACCATATCAAAGATGGAGGACCACAT 854

Qy        180 GlyPheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArg 199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 GGATTTACACAACAGCCTTTGCGGCCCAATTTAGTGGAAACTAGTTGTCCCAATATACGA 914

Qy        200 MetAspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAsp 219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        915 ATGGATCCAAAATTATGCCCTGCTGATCCAGACTGGATAGCTTTCATTCACAGCAATGAT 974

Qy        220 IleTrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGlu 239
              ||||||||||||||:::|||||||||||||||||||:::|||||||||||||||||||||
Db        975 ATTTGGATATCAAACCTCGTAACCAGGGAAGAGCGGAGGATCACATACGTACACAATGAG 1034

Qy        240 LeuAlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGlu 259
              ||||||||||||||||||||    ||||||||||||||||||||||||||||||||||||
Db       1035 CTAGCCAACATGGAAGAGGATCCCAGATCAGCTGGTGTTGCCACCTTTGTTCTTCAAGAA 1094

Qy        260 GluPheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGly 279
              ||||||||||||||||||||||||||||||||||||:::|||||||    ||||||||||||
Db       1095 GAATTTGACAGATACTCTGGCTACTGGTGGTGTCCCCAAGCGGAAAGAACTCCTAGTGGT 1154

Qy        280 GlyLysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHis 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1155 GGTAAAATTCTTAGAATTCTCTATGAAGAAAATGATGAATCTGAGGTGGAGATTATTCAT 1214

Qy        300 ValThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGly 319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1215 GTTACGTCCCCCATGTTGGAAACAAGGAGGGCAGATTCCTTTCGTTATCCCAAAACAGGC 1274

Qy        320 ThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArg 339
              ||||||||||||||||||||||||||||||||||||:::::||||||   |||
Db       1275 ACAGCAAACCCAAAGGTCACTTTCAAGATGTCGGAGATTGTTGTTGATGCTGCAGGAGGG 1334

Qy        340 IleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyVal 359
              ||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db       1335 ATTATAGATGTCATAGATAAAGAACTGGTTCAACCTTTCGAGATTCTGTTTGAGGGAGTT 1394

Qy        360 GluTyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeu 379
              ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db       1395 GAATATATTGCCAGAGCCGGATGGACTCCAGAGGGAAAACATGCCTGGTCCATCCTACTA 1454

Qy        380 AspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProVal 399
              ||||||||||||||    ||||||||||||||||||||||||||||||||||||||||||
Db       1455 GACCGTTCCCAGACTCACCTGCAGATAGTTCTGATCTCCCCTGAGTTATTCATCCCAGTA 1514
```

```
Qy        400 GluAspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrPro 419
              |||||||||    |||:::||||||||||||||||||||||||||||||||||||||||||
Db       1515 GAAGATGATGCCATGGACAGACAGAGACTTATAGAGTCAGTTCCTGACTCTGTGACACCA 1574

Qy        420 LeuIleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisVal 439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1575 CTGATCATCTATGAAGAAACAACAGACATCTGGATAAATATCCACGATATTTTTCATGTT 1634

Qy        440 PheProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGly 459
              |||||||||:::||||||:::|||||||||||||||||||||||||||||||||||||||
Db       1635 TTTCCTCAAACTCATGAAGATGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGT 1694

Qy        460 PheArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSer 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1695 TTTCGTCATCTGTATAAAATCACATCCATTTTAAAGGAGAGCAAATATAAACGGTCCAGT 1754

Qy        480 GlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThr 499
              ||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db       1755 GGTGGACTACCTGCCCCAAGTGATTTCAAGTGTCCTATCAAAGAAGAAATAACAATTACC 1814

Qy        500 SerGlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArg 519
              ||||||||||||||||||||||||||||||||||||||||||   ||||||||    |||
Db       1815 AGTGGTGAATGGGAAGTACTTGGCCGGCATGGATCTAATATCTGGGTTGATGAAGCCAGA 1874

Qy        520 ArgLeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValVal 539
              :::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1875 AAGCTGGTGTACTTTGAAGGCACCAAAGACTCTCCTTTGGAGCATCACCTGTACGTGACC 1934

Qy        540 SerTyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCys 559
              ||||||     |||||||||||||||     ||||||||||||||||||||||||||||||
Db       1935 AGTTATGCAAACCCTGGAGAAGTGGTGAGGCTGACTGACCGTGGCTACTCACACTCCTGC 1994

Qy        560 CysIleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHis 579
              |||:::|||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db       1995 TGCCTCAGCCGGCATTGTGACTTCTTCATAAGTAAGTACAGCAACCAGAAGAATCCACAC 2054

Qy        580 CysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGlu 599
              ||||||||||||||||||||||||||||||||||||||||||      ||||||||||||
Db       2055 TGTGTGTCCCTCTACAAACTCTCAAGTCCTGAGGATGACCCAGTTCATAAAACAAAGGAA 2114

Qy        600 PheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIle 619
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2115 TTTTGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTACACCCCTCCAGAAATT 2174

Qy        620 PheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAsp 639
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2175 TTTTCTTTTGAAAGTACTACTGGATTTACACTGTATGGAATGTTGTATAAGCCTCATGAC 2234

Qy        640 LeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGln 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2235 CTACAACCTGGAAAGAAATACCCCACTGTGTTATTCATATATGGTGGTCCCCAGGTGCAG 2294

Qy        660 LeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeu 679
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2295 CTGGTGAACAATCGGTTTAAAGGAGTCAAGTATTTCCGCCTGAACACCCTGGCCTCCCTG 2354

Qy        680 GlyTyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGlu 699
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2355 GGTTATGTGGTTGTGGTGATAGACAACAGGGGATCCTGTCACCGAGGACTTAAATTTGAA 2414

Qy        700 GlyAlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGln 719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2415 GGCGCCTTTAAATATAAAATGGGTCAAATAGAAATCGATGATCAAGTGGAAGGACTCCAG 2474

Qy        720 TyrLeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSer 739
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db       2475 TACCTAGCATCTCAGTATGACTTCATTGACTTGGATCGAGTGGGCATCCACGGCTGGTCC 2534

Qy        740 TyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAla 759
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2535 TATGGTGGCTACCTCTCCCTGATGGCATTAATGCAGAGGTCGGATATCTTCCGGGTGGCT 2594

Qy        760 IleAlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyr 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2595 ATTGCTGGGGCCCCAGTCACCCTGTGGATCTTCTATGATACAGGATACACGGAGCGCTAT 2654

Qy        780 MetGlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAla 799
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2655 ATGGGTCACCCTGACCAGAATGAACAGGGCTACTACTTAGGATCTGTGGCCATGCAAGCG 2714

Qy        800 GluLysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsn 819
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2715 GAGAAGTTCCCCTCAGAACCAAACCGGTTACTCCTCTTGCATGGGTTCTTGGATGAGAAT 2774

Qy        820 ValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyr 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2775 GTTCACTTTGCACACACCAGTATATTGCTGAGTTTTTTAGTGAGGGCTGGAAAGCCATAT 2834

Qy        840 AspLeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHis 859
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2835 GACTTACAGATCTATCCTCAGGAGAGGCACAGCATCAGAGTTCCTGAGTCTGGAGAACAC 2894

Qy        860 TyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeu 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2895 TATGAACTGCACCTGCTCCACTACCTTCAGGAGAACCTTGGATCGCGTATCGCTGCTCTG 2954

Qy        880 LysValIle 882
              |||||||||
Db       2955 AAAGTGATA 2963

RESULT 10
BC030688
LOCUS       BC030688                3130 bp    mRNA    linear   PRI 30-MAY-2002
DEFINITION  Homo sapiens, Similar to dipeptidylpeptidase 8, clone MGC:26191
            IMAGE:4822550, mRNA, complete cds.
ACCESSION   BC030688
VERSION     BC030688.1  GI:21265132
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3130)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 33 Row: d Column: 5
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 18450277.
FEATURES             Location/Qualifiers
     source          1..3130
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:26191 IMAGE:4822550"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
```

```
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
     CDS             372. .2867
                     /codon_start=1
                     /product="Similar to dipeptidylpeptidase 8"
                     /protein_id="AAH30688.1"
                     /db_xref="GI:21265133"
                     /translation="MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQ
                     LKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEI
                     PKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSG
                     TFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSN
                     DIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETT
                     PSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIMI
                     DAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLIS
                     PELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSHEEEIEF
                     IFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGR
                     HGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCD
                     FFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFES
                     TTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVQLVNNRFKGVKYFRLNTLASLGYV
                     VVVIDNRGSCHRGLKFEGAFKYKMVAIAGAPVTLWIFYDTGYTERYMGHPDQNEQGYY
                     LGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYDLQIYPQERH
                     SIRVPESGEHYELHLLHYLQENLGSRIAALKVI"
BASE COUNT      955 a    626 c    717 g    832 t
ORIGIN

Alignment Scores:
Pred. No.:                0              Length:          3130
Score:                    4397.50        Matches:         831
Percent Similarity:       94.22%         Conservative:    0
Best Local Similarity:    94.22%         Mismatches:      0
Query Match:              93.56%         Indels:          51
DB:                       9              Gaps:            1

US-10-070-464-1 (1-882) x BC030688 (1-3130)

Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 431

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 491

Qy         41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 551

Qy         61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        552 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 611

Qy         81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 671

Qy        101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        672 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 731

Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        732 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 791

Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        792 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 851

Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 911

Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        912 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 971

Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        972 GATCCAAAATTATGCCCTGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 1031

Qy        221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1032 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 1091

Qy        241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1092 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 1151

Qy        261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1152 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 1211

Qy        281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1212 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 1271

Qy        301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1272 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 1331

Qy        321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1332 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1391

Qy        341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1392 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1451

Qy        361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1452 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1511

Qy        381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1512 CGCTCCCAGACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1571

Qy        401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1572 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1631

Qy        421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1632 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1691

Qy        441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1692 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 1751

Qy        461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1752 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1811

Qy        481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1812 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1871

Qy        501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1872 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1931

Qy        521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1932 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1991

Qy        541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1992 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 2051
```

```
Qy        561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2052 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 2111

Qy        581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2112 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 2171

Qy        601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2172 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 2231

Qy        621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2232 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 2291

Qy        641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2292 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTG 2351

Qy        661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2352 GTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGT 2411

Qy        681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2412 TATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGC 2471

Qy        701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
              ||||||||||||||||||
Db       2472 GCCTTTAAATATAAAATG---------------------------------------- 2489

Qy        721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740

Db       2489 ---------------------------------------------------------- 2489

Qy        741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
                                                        |||||||||
Db       2490 ----------------------------------------------------GTTGCTATT 2498

Qy        761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2499 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 2558

Qy        781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2559 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 2618

Qy        801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2619 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 2678

Qy        821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2679 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 2738

Qy        841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2739 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATTAT 2798

Qy        861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2799 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 2858

Qy        881 ValIle 882
              ||||||
Db       2859 GTGATA 2864

RESULT 11
AX608745
LOCUS       AX608745               4685 bp    DNA     linear   PAT 17-FEB-2003
DEFINITION  Sequence 22 from Patent WO0231134.
ACCESSION   AX608745
VERSION     AX608745.1  GI:28404308
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Qi,S., Akinsanya,K.O., Riviere,P.J. and Junien,J.L.
  TITLE     Novel serine protease genes related to dppiv
  JOURNAL   Patent: WO 0231134-A 22 18-APR-2002;
            Ferring BV (NL)
FEATURES             Location/Qualifiers
     source          1. .4685
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT     1430 a    853 c    991 g   1411 t
ORIGIN

Alignment Scores:
Pred. No.:                0              Length:          4685
Score:                    4385.50        Matches:         834
Percent Similarity:       94.56%         Conservative:    0
Best Local Similarity:    94.56%         Mismatches:      1
Query Match:              93.31%         Indels:          48
DB:                       6              Gaps:            1

US-10-070-464-1 (1-882) x AX608745 (1-4685)

Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 273

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 333

Qy         41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 393

Qy         61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        394 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 453

Qy         81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 513

Qy        101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 573

Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 633

Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 693

Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        694 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 753

Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        754 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 813

Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        814 GATCCAAAATTATGCCCTGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 873
```

```
Qy      221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      874 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 933

Qy      241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      934 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 993

Qy      261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      994 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 1053

Qy      281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1054 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 1113

Qy      301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1114 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 1173

Qy      321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1174 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1233

Qy      341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1234 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1293

Qy      361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1294 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1353

Qy      381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1354 CGCTCCCAGACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1413

Qy      401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1414 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1473

Qy      421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1474 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1533

Qy      441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1534 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 1593

Qy      461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1594 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1653

Qy      481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1654 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1713

Qy      501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1714 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1773

Qy      521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1774 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1833

Qy      541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1834 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 1893

Qy      561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1894 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 1953

Qy      581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1954 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 2013

Qy      601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
            ||||||||||||||||||||
Db     2014 TGGGCCACCATTTTGGATTCAGT---------------------------------- 2036

Qy      621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640

Db     2036 ---------------------------------------------------------- 2036

Qy      641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
                                                     ||||||||||||||||
Db     2037 ----------------------------------------------CCTCAGGTGCAGTTG 2051

Qy      661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2052 GTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGT 2111

Qy      681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2112 TATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGC 2171

Qy      701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2172 GCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATAT 2231

Qy      721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2232 CTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTAT 2291

Qy      741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2292 GGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATT 2351

Qy      761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2352 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 2411

Qy      781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2412 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 2471

Qy      801 LysProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2472 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 2531

Qy      821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2532 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 2591

Qy      841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2592 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATTAT 2651

Qy      861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2652 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 2711

Qy      881 ValIle 882
            ||||||
Db     2712 GTGATA 2717
```

```
RESULT 12
AX608743
LOCUS       AX608743               4676 bp    DNA     linear   PAT 17-FEB-2003
DEFINITION  Sequence 20 from Patent WO0231134.
ACCESSION   AX608743
VERSION     AX608743.1  GI:28404307
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1
  AUTHORS  Qi,S., Akinsanya,K.O., Riviere,P.J. and Junien,J.L.
  TITLE    Novel serine protease genes related to dppiv
  JOURNAL  Patent: WO 0231134-A 20 18-APR-2002;
           Ferring BV (NL)
FEATURES             Location/Qualifiers
     source          1. .4676
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT     1424 a    859 c    979 g   1414 t
ORIGIN
```

Alignment Scores:

Pred. No.:	0	Length:	4676
Score:	4385.00	Matches:	831
Percent Similarity:	94.22%	Conservative:	0
Best Local Similarity:	94.22%	Mismatches:	1
Query Match:	93.30%	Indels:	51
DB:	6	Gaps:	1

US-10-070-464-1 (1-882) x AX608743 (1-4676)

```
Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 273

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 333

Qy         41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 393

Qy         61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        394 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 453

Qy         81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 513

Qy        101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 573

Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 633

Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 693

Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        694 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 753

Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        754 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 813

Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        814 GATCCAAAATTATGCCCTGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 873

Qy        221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        874 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 933

Qy        241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        934 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 993

Qy        261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        994 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 1053

Qy        281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1054 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 1113

Qy        301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 1173

Qy        321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1174 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1233

Qy        341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1234 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1293

Qy        361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1294 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1353

Qy        381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1354 CGCTCCCAGACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1413

Qy        401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1414 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1473

Qy        421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1474 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1533

Qy        441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1534 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 1593

Qy        461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1594 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1653

Qy        481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1654 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1713

Qy        501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1714 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1773

Qy        521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1774 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1833

Qy        541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1834 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 1893

Qy        561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1894 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 1953

Qy        581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1954 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 2013

Qy        601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
```

```
Db       2014 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 2073

Qy        621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640

Db       2074 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 2133

Qy        641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660

Db       2134 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCG---------- 2180

Qy        661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680

Db       2180 ---------------------------------------------------------- 2180

Qy        681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700

Db       2180 ---------------------------------------------------------- 2180

Qy        701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720

Db       2181 -----------------GGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATAT 2222

Qy        721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740

Db       2223 CTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTAT 2282

Qy        741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760

Db       2283 GGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATT 2342

Qy        761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780

Db       2343 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 2402

Qy        781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800

Db       2403 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 2462

Qy        801 LysProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820

Db       2463 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 2522

Qy        821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840

Db       2523 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 2582

Qy        841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860

Db       2583 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATTAT 2642

Qy        861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880

Db       2643 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 2702

Qy        881 ValIle 882

Db       2703 GTGATA 2708
```

```
RESULT 13
AX405770
LOCUS       AX405770               2842 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 185 from Patent WO0222660.
ACCESSION   AX405770
VERSION     AX405770.1  GI:21438979
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tang,Y.T., Liu,C., Zhou,P., Asundi,V., Zhang,J., Zhao,Q.A., Ren,F.,
            Xue,A.J., Yang,Y., Wehrman,T. and Drmanac,R.T.
  TITLE     Novel nucleic acids and polypeptides
  JOURNAL   Patent: WO 0222660-A 185 21-MAR-2002;
            HYSEQ, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .2842
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CDS             234. .2582
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD34877.1"
                     /db_xref="GI:21438980"
                     /translation="MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQ
                     LKKLLADTRKYHGYMMAKAPHDFMPVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEI
                     PKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSG
                     TFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSN
                     DIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETT
                     PSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIMI
                     DAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLIS
                     PELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSHEEEIEF
                     IFASECKTGFRHLYKITSILKESKYKRSSGGLPAPSDFKCPIKEEIAITSGEWEVLGR
                     HGSNIQVDEVRRLVYFEGTKDSPLEHHLYVVSYVNPGEVTRLTDRGYSHSCCISQHCD
                     FFISKYSNQKNPHCVSLYKLSSPEDDPTCKTKEFWATILDSAGPLPDYTPPEIFSFES
                     TTGFTLYGMLYKPHDLQPGKKYPTVLFIYGGPQVAIAGAPVTLWIFYDTGYTERYMGH
                     PDQNEQGYYLGSVAMQAEKFPSEPNRLLLLHGFLDENVHFAHTSILLSFLVRAGKPYD
                     LQIYPQERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI"
BASE COUNT      857 a    592 c    635 g    758 t
ORIGIN
```

Alignment Scores:

Pred. No.:	0	Length:	2842
Score:	4118.00	Matches:	782
Percent Similarity:	88.66%	Conservative:	0
Best Local Similarity:	88.66%	Mismatches:	0
Query Match:	87.62%	Indels:	100
DB:	6	Gaps:	1

US-10-070-464-1 (1-882) x AX405770 (1-2842)

```
Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20

Db        234 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 293

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40

Db        294 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 353

Qy         41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60

Db        354 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 413

Qy         61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80

Db        414 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 473

Qy         81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100

Db        474 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 533

Qy        101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120

Db        534 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 593

Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140

Db        594 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 653

Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160

Db        654 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 713

Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
```

```
Db        714 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 773
Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
Db        774 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 833
Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
Db        834 GATCCAAAATTATGCCCTGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 893
Qy        221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
Db        894 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 953
Qy        241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
Db        954 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 1013
Qy        261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
Db       1014 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 1073
Qy        281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
Db       1074 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 1133
Qy        301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
Db       1134 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 1193
Qy        321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
Db       1194 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1253
Qy        341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
Db       1254 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1313
Qy        361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
Db       1314 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1373
Qy        381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
Db       1374 CGCTCCCAGACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1433
Qy        401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
Db       1434 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1493
Qy        421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
Db       1494 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1553
Qy        441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
Db       1554 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 1613
Qy        461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
Db       1614 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1673
Qy        481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
Db       1674 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1733
Qy        501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
Db       1734 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1793
Qy        521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
Db       1794 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1853
Qy        541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
Db       1854 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 1913
Qy        561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
Db       1914 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 1973
Qy        581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
Db       1974 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 2033
Qy        601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
Db       2034 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 2093
Qy        621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db       2094 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 2153
Qy        641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
Db       2154 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAG--------- 2204
Qy        661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db       2204 ---------------------------------------------------------- 2204
Qy        681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
Db       2204 ---------------------------------------------------------- 2204
Qy        701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
Db       2204 ---------------------------------------------------------- 2204
Qy        721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740
Db       2204 ---------------------------------------------------------- 2204
Qy        741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
Db       2205 ----------------------------------------------------GTTGCTATT 2213
Qy        761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
Db       2214 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 2273
Qy        781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
Db       2274 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 2333
Qy        801 LysProProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
Db       2334 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 2393
Qy        821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
Db       2394 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 2453
Qy        841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
Db       2454 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATTAT 2513
Qy        861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
Db       2514 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 2573
Qy        881 ValIle 882
Db       2574 GTGATA 2579
```

```
RESULT 14
AX608731
LOCUS       AX608731                4523 bp    DNA     linear   PAT 17-FEB-2003
DEFINITION  Sequence 8 from Patent WO0231134.
ACCESSION   AX608731
VERSION     AX608731.1  GI:28404301
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Qi,S., Akinsanya,K.O., Riviere,P.J. and Junien,J.L.
  TITLE     Novel serine protease genes related to dppiv
  JOURNAL   Patent: WO 0231134-A 8 18-APR-2002;
            Ferring BV (NL)
FEATURES             Location/Qualifiers
     source          1. .4523
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT     1384 a    828 c    940 g   1371 t
ORIGIN

Alignment Scores:
Pred. No.:              0              Length:         4523
Score:                  4092.50        Matches:        780
Percent Similarity:     88.44%         Conservative:   0
Best Local Similarity:  88.44%         Mismatches:     1
Query Match:            87.07%         Indels:         102
DB:                     6              Gaps:           1

US-10-070-464-1 (1-882) x AX608731 (1-4523)

Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 273

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 333

Qy         41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 393

Qy         61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        394 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 453

Qy         81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 513

Qy        101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 573

Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 633

Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 693

Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        694 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 753

Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        754 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 813

Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        814 GATCCAAAATTATGCCCTGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 873

Qy        221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        874 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 933

Qy        241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        934 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 993

Qy        261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        994 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 1053

Qy        281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1054 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 1113

Qy        301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 1173

Qy        321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1174 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1233

Qy        341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1234 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1293

Qy        361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1294 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1353

Qy        381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1354 CGCTCCCAGACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1413

Qy        401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1414 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1473

Qy        421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1474 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1533

Qy        441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1534 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 1593

Qy        461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1594 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1653

Qy        481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1654 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1713

Qy        501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1714 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1773

Qy        521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1774 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1833

Qy        541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1834 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 1893
```

```
Qy        561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1894 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 1953

Qy        581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1954 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 2013

Qy        601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2014 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 2073

Qy        621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2074 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 2133

Qy        641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
              ||||||||||||||||||||||||||||||||||||||
Db       2134 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGT---------- 2178

Qy        661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680

Db       2178 ---------------------------------------------------------- 2178

Qy        681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700

Db       2178 ---------------------------------------------------------- 2178

Qy        701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720

Db       2178 ---------------------------------------------------------- 2178

Qy        721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740

Db       2178 ---------------------------------------------------------- 2178

Qy        741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
                                                      |||   ||||||
Db       2179 ----------------------------------------------CGGTT-GCTATT 2189

Qy        761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2190 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 2249

Qy        781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2250 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 2309

Qy        801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2310 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 2369

Qy        821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2370 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 2429

Qy        841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2430 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATTAT 2489

Qy        861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2490 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 2549

Qy        881 ValIle 882
              ||||||
Db       2550 GTGATA 2555

RESULT 15
AK000290
LOCUS       AK000290               2778 bp    mRNA    linear   PRI 22-FEB-2000
DEFINITION  Homo sapiens cDNA FLJ20283 fis, clone HEP04088.
ACCESSION   AK000290
VERSION     AK000290.1  GI:7020272
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
            Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,
            Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
            Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2778)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Deptment of Virology; Shirokane-dai, 4-6-1,
            Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna1@ims.u-tokyo.ac.jp,
            Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Departent of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1. .2778
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HEP04088"
                     /cell_line="HepG2"
                     /cell_type="hepatoma"
                     /clone_lib="HEP"
                     /note="cloning vector pME18SFL3"
     CDS             163. .1572
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAA91059.1"
                     /db_xref="GI:7020273"
                     /translation="MAAAMETEQLGVEIFETADCEENIESQDRPKLEPFYVERYSWSQ
                     LKKLLADTRKYHGYMMAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEI
                     PKTINRAAVLMLSWKPLLDLFQATLDYGMYSREEELLRERKRIGTVGIASYDYHQGSG
                     TFLFQAGSGIYHVKDGGPQGFTQQPLRPNLVETSCPNIRMDPKLCPADPDWIAFIHSN
                     DIWISNIVTREERRLTYVHNELANMEEDARSAGVATFVLQEEFDRYSGYWWCPKAETT
                     PSGGKILRILYEENDESEVEIIHVTSPMLETRRADSFRYPKTGTANPKVTFKMSEIMI
                     DAEGRIIDVIDKELIQPFEILFEGVEYIARAGWTPEGKYAWSILLDRSQTRLQIVLIS
                     PELFIPVEDDVMERQRLIESVPDSVTPLIIYEETTDIWINIHDIFHVFPQSHEEEIEF
                     IFASECKTGFRHLYKITSI"
BASE COUNT      856 a    562 c    615 g    745 t
ORIGIN

Alignment Scores:
Pred. No.:                0              Length:          2778
Score:                    4091.00        Matches:         781
Percent Similarity:       88.55%         Conservative:    0
Best Local Similarity:    88.55%         Mismatches:      1
Query Match:              87.04%         Indels:          101
DB:                       9              Gaps:            1

US-10-070-464-1 (1-882) x AK000290 (1-2778)

Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        163 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 222

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        223 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 282
```

```
Qy        41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
Db       283 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 342

Qy        61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
Db       343 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 402

Qy        81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
Db       403 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 462

Qy       101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
Db       463 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 522

Qy       121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
Db       523 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 582

Qy       141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
Db       583 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 642

Qy       161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
Db       643 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 702

Qy       181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
Db       703 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 762

Qy       201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
Db       763 GATCCAAAATTATGCCCTGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 822

Qy       221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
Db       823 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 882

Qy       241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
Db       883 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 942

Qy       261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
Db       943 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 1002

Qy       281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
Db      1003 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 1062

Qy       301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
Db      1063 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 1122

Qy       321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
Db      1123 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1182

Qy       341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
Db      1183 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1242

Qy       361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
Db      1243 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1302

Qy       381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
Db      1303 CGCTCCCAGACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1362

Qy       401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
Db      1363 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1422

Qy       421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
Db      1423 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1482

Qy       441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
Db      1483 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 1542

Qy       461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
Db      1543 CGTCATTTATACAAAATTACATCTAT-TTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1601

Qy       481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
Db      1602 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1661

Qy       501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
Db      1662 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1721

Qy       521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
Db      1722 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1781

Qy       541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
Db      1782 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 1841

Qy       561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
Db      1842 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 1901

Qy       581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
Db      1902 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 1961

Qy       601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
Db      1962 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 2021

Qy       621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db      2022 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 2081

Qy       641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
Db      2082 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAG--------- 2132

Qy       661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db      2132 ------------------------------------------------------------ 2132

Qy       681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
Db      2132 ------------------------------------------------------------ 2132

Qy       701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
Db      2132 ------------------------------------------------------------ 2132

Qy       721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740
Db      2132 ------------------------------------------------------------ 2132

Qy       741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
Db      2133 ----------------------------------------------------GTTGCTATT 2141

Qy       761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           2142 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 2201

Qy            781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           2202 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 2261

Qy            801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           2262 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 2321

Qy            821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           2322 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 2381

Qy            841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           2382 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATTAT 2441

Qy            861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           2442 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 2501

Qy            881 ValIle 882
                ||||||
Db           2502 GTGATA 2507
```

Search completed: October 16, 2003, 00:03:57
Job time : 9759.41 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 15, 2003, 17:14:56 ; Search time 648.054 Seconds
(without alignments)
3673.932 Million cell updates/sec

Title: US-10-070-464-1
Perfect score: 4700
Sequence: 1 MAAAMETEQLGVEIFETADC.....HLLHYLQENLGSRIAALKVI 882

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10070464/runat_15102003_113552_24799/app_query.fasta_1.2652
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10070464_@CGN_1_1_0_@runat_15102003_113552_24799 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4700	100.0	2649	25	ABX12255	cDNA encoding huma
2	4700	100.0	2671	24	ABK83322	cDNA encoding huma
3	4700	100.0	3106	24	ABK12892	Human protease PRT
4	4700	100.0	3120	22	AAC85694	Nucleotide sequenc
5	4700	100.0	3120	24	AAD38956	Human dipeptidyl p
6	4700	100.0	3143	24	AAH99934	cDNA encoding 2195
7	4695	99.9	2643	24	AAH99935	Coding sequence of
8	4680	99.6	4829	24	ABK83327	cDNA encoding huma
9	4385.5	93.3	4685	24	ABK83332	cDNA encoding huma
10	4385	93.3	4676	24	ABK83331	cDNA encoding huma
11	4118	87.6	2842	24	ABN59774	Novel human coding
12	4092.5	87.1	4523	24	ABK83325	cDNA encoding huma
13	3970.5	84.5	2510	24	AAD23843	Human protease PRT
14	3771	80.2	2668	24	ABN59775	Novel human coding
15	3661.5	77.9	4309	24	ABK83328	cDNA encoding huma
16	3364.5	71.6	2161	22	AAH15009	Human cDNA sequenc
17	2870	61.1	2617	24	ABK83323	cDNA encoding huma
18	2870	61.1	3716	24	ABQ75955	Human PMMM encodin
19	2870	61.1	4219	24	ABK83335	cDNA encoding huma
20	2870	61.1	4302	24	ABK83333	cDNA encoding huma
21	2863	60.9	3024	24	AAD38954	Human dipeptidyl p
22	2835	60.3	2495	24	AAD38957	Human dipeptidyl p
23	2833	60.3	3287	24	AAD38955	Alternative versio
24	2820.5	60.0	4180	24	ABK83339	cDNA encoding huma
25	2820.5	60.0	4263	24	ABK83338	cDNA encoding huma
26	2801	59.6	1821	24	ABV76411	Dipeptidyl peptida
27	2763	58.8	2751	24	AAD38311	Murine dipeptidyl
28	2649	56.4	4076	24	ABK83337	cDNA encoding huma
29	2649	56.4	4159	24	ABK83336	cDNA encoding huma
30	2638	56.1	2801	22	AAI57896	Human polynucleoti
31	2599.5	55.3	4037	24	ABK83341	cDNA encoding huma
32	2599.5	55.3	4120	24	ABK83340	cDNA encoding huma
33	2476.5	52.7	3262	22	AAI57880	Human polynucleoti
34	2422	51.5	1669	22	AAC85696	Nucleotide sequenc
35	2226.5	47.4	2982	22	AAI59666	Human polynucleoti
36	1914.5	40.7	2461	21	AAC75835	Human ORFX ORF1390
37	1878.5	40.0	2952	24	ABK69090	DNA encoding human
38	1874.5	39.9	3047	24	ABK69113	DNA encoding human
39	1836.5	39.1	1083	22	AAC85697	Nucleotide sequenc
40	1645.5	35.0	1197	22	AAC85695	Nucleotide sequenc
41	1644.5	35.0	2027	21	AAC77137	Human ORFX ORF2692
42	1599.5	34.0	3713	23	ABL10425	Drosophila melanog
43	1599.5	34.0	3783	23	ABL06641	Drosophila melanog
44	1400	29.8	2079	21	AAA37672	Human peptidase, H
45	1400	29.8	2079	25	ABX63843	Human cDNA #843 di

ALIGNMENTS

RESULT 1
ABX12255
ID ABX12255 standard; cDNA; 2649 BP.
XX
AC ABX12255;
XX
DT 19-MAY-2003 (first entry)
XX
DE cDNA encoding human serine protease HIPHUM46.
XX
KW Human; ss; gene; HIPHUM46; serine protease; gene therapy; osteoarthritis;
KW serine protease activity modulation; dipeptidyl peptidase activity;
KW musculoskeletal disease; Hepatitis B virus infection; myotonic dystrophy;
KW Alzheimer's disease; parasupranuclear palsy; Huntington's disease;
KW amyotrophic lateral sclerosis; malabsorption syndrome; lung disease;
KW irritable bowel syndrome; type I diabetes; faecal incontinence; anaemia;
KW haemorrhoid; proctitis; rectal polyp; small bowel tumour; dyslexia;

KW colorectal tumour; ceroid lipofuscinosis; allergic encephalomyelitis;
KW multiple sclerosis; chromosome 15q21-q22.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..2649
FT /*tag= a
FT /product= "HIPHUM46"
FT /note= "Serine protease"
XX
PN GB2374869-A.
XX
PD 30-OCT-2002.
XX
PF 22-JAN-2002; 2002GB-0001404.
XX
PR 23-JAN-2001; 2001GB-0001760.
XX
PA (GLAX) GLAXO GROUP LTD.
XX
PI Edbrooke MR, Lewis AP;
XX
DR WPI; 2003-150703/15.
DR P-PSDB; ABU07720.
XX
PT Identifying modulators of serine protease activity useful for treating
PT musculoskeletal diseases, by contacting cell expressing a novel serine
PT protease polypeptide with a compound and monitoring serine protease
PT activity -
XX
PS Claim 12; Page 22-26; 38pp; English.
XX
CC The invention relates to a method of identifying a substance that
CC modulates serine protease activity, comprising contacting a cell such as
CC a neuronal cell, lung cell, intestinal cell or a cell infected with a
CC virus, expressing a serine protease polypeptide (HIPHUM 46), or its
CC variant having dipeptidyl peptidase activity, or a serine protease
CC isolated from the cell with a test substance and monitoring for serine
CC protease activity. The method is useful for identifying a substance that
CC modulates serine protease activity. A modulator of the serine protease is
CC useful in the manufacture of a medicament for treatment or prophylaxis of
CC a musculoskeletal disease e.g. osteoarthritis, Hepatitis B virus
CC infection, Alzheimer's disease, parasupranuclear palsy, myotonic
CC dystrophy, Huntington's disease or amyotrophic lateral sclerosis.
CC Additional disease that may be treated using modulators of the serine
CC protease include malabsorption syndromes, irritable bowel syndrome, lung
CC disease, type I diabetes, faecal incontinence, haemorrhoids, proctitis,
CC rectal polyps, small bowel tumours, colorectal tumours, anaemia,
CC dyslexia, ceroid lipofuscinosis, allergic encephalomyelitis, and multiple
CC sclerosis. The present sequence represents cDNA encoding the human serine
CC protease HIPHUM46 which is located on chromosome 15q21-q22.
XX
SQ Sequence 2649 BP; 803 A; 514 C; 585 G; 747 T; 0 other;

Alignment Scores:

Pred. No.:	0	Length:	2649
Score:	4700.00	Matches:	882
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	25	Gaps:	0

US-10-070-464-1 (1-882) x ABX12255 (1-2649)

```
Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 60

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 120

Qy         41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 180

Qy         61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 240

Qy         81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 300

Qy        101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 360

Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 420

Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 480

Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 540

Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 600

Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GATCCAAAATTATGCCCTGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 660

Qy        221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 720

Qy        241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 780

Qy        261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 840

Qy        281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 900

Qy        301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 960

Qy        321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1020

Qy        341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1080

Qy        361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1140

Qy        381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CGCTCCCAGACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1200

Qy        401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
```

```
Db       1201 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1260

Qy        421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440

Db       1261 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1320

Qy        441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460

Db       1321 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTGCCTCTGAATGCAAAACAGGTTTC 1380

Qy        461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480

Db       1381 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1440

Qy        481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500

Db       1441 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1500

Qy        501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520

Db       1501 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1560

Qy        521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540

Db       1561 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1620

Qy        541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560

Db       1621 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 1680

Qy        561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580

Db       1681 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 1740

Qy        581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600

Db       1741 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 1800

Qy        601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620

Db       1801 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 1860

Qy        621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640

Db       1861 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 1920

Qy        641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660

Db       1921 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTG 1980

Qy        661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680

Db       1981 GTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGT 2040

Qy        681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700

Db       2041 TATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGC 2100

Qy        701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720

Db       2101 GCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATAT 2160

Qy        721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740

Db       2161 CTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTAT 2220

Qy        741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760

Db       2221 GGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATT 2280

Qy        761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780

Db       2281 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 2340

Qy        781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800

Db       2341 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 2400

Qy        801 LysPheProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820

Db       2401 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 2460

Qy        821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840

Db       2461 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 2520

Qy        841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860

Db       2521 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCTGGAGAACATTAT 2580

Qy        861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880

Db       2581 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 2640

Qy        881 ValIle 882

Db       2641 GTGATA 2646
```

RESULT 2
ABK83322
ID ABK83322 standard; cDNA; 2671 BP.
XX
AC ABK83322;
XX
DT 12-AUG-2002 (first entry)
XX
DE cDNA encoding human DPPIV related serine protease DPRP-1.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI; 2002-444178/47.
DR P-PSDB; ABG61591.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT -
XX
PS Claim 1; Page 53-54; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly

CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABK83322-ABK83343 encode human DPRP proteins.
XX
SQ Sequence 2671 BP; 805 A; 524 C; 594 G; 748 T; 0 other;

Alignment Scores:

Pred. No.:	0	Length:	2671
Score:	4700.00	Matches:	882
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	24	Gaps:	0

US-10-070-464-1 (1-882) x ABK83322 (1-2671)

```
Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          8 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 67

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 127

Qy         41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        128 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 187

Qy         61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 247

Qy         81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 307

Qy        101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 367

Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 427

Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 487

Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        488 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 547

Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        548 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 607

Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        608 GATCCAAAATTATGCCCTGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 667

Qy        221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        668 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 727

Qy        241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        728 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 787

Qy        261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        788 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 847

Qy        281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 907

Qy        301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 967

Qy        321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1027

Qy        341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1087

Qy        361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1147

Qy        381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 CGCTCCCAGACTCGCCTGCAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1207

Qy        401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1267

Qy        421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1268 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1327

Qy        441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1328 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 1387

Qy        461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1388 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1447

Qy        481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1448 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1507

Qy        501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1508 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1567

Qy        521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1568 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1627

Qy        541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1628 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 1687

Qy        561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1688 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 1747

Qy        581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1748 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 1807

Qy        601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1808 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 1867
```

```
Qy        621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1868 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 1927

Qy        641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1928 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTG 1987

Qy        661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1988 GTGAATAATCGATTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGT 2047

Qy        681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2048 TATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGC 2107

Qy        701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2108 GCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATAT 2167

Qy        721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2168 CTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTAT 2227

Qy        741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2228 GGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATT 2287

Qy        761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2288 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 2347

Qy        781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2348 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 2407

Qy        801 LysProProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2408 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 2467

Qy        821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2468 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 2527

Qy        841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2528 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATTAT 2587

Qy        861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2588 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 2647

Qy        881 ValIle 882
              ||||||
Db       2648 GTGATA 2653
```

RESULT 3
ABK12892
ID ABK12892 standard; cDNA; 3106 BP.
XX
AC ABK12892;
XX
DT 09-APR-2002 (first entry)
XX
DE Human protease PRTS-9 cDNA sequence.
XX
KW Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KW cell proliferative disorder; developmental disorder; epilepsy;
KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KW reproductive disorder; endometriosis; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   203..2851
FT   /*tag= a
FT   /product= "Human protease PRTS-9"
XX
PN WO200198468-A2.
XX
PD 27-DEC-2001.
XX
PF 13-JUN-2001; 2001WO-US19178.
XX
PR 16-JUN-2000; 2000US-212336P.
PR 22-JUN-2000; 2000US-213955P.
PR 29-JUN-2000; 2000US-215396P.
PR 07-JUL-2000; 2000US-216821P.
PR 14-JUL-2000; 2000US-218946P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM;
PI Delegeane AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
PI Walia NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT;
PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
PI Kallick DA;
XX
DR WPI; 2002-090437/12.
DR P-PSDB; AAU74749.
XX
PT Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
PT proliferative (e.g. cancer)disorders -
XX
PS Claim 5; Page 166-167; 177pp; English.
XX
CC The present invention relates to twenty one new human proteases,
CC referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
CC polypeptides of the invention are useful in the diagnosis, treatment and
CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
CC myocardial infarction, autoimmune/inflammatory e.g. acquired
CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
CC endometriosis disorders. Numerous other examples of each disorder are
CC given in the specification. The present nucleic acid sequence encodes
CC the human protease PRTS-9 protein of the invention.
XX
SQ Sequence 3106 BP; 928 A; 633 C; 704 G; 841 T; 0 other;

Alignment Scores:

Pred. No.:	0	Length:	3106
Score:	4700.00	Matches:	882
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	24	Gaps:	0

US-10-070-464-1 (1-882) x ABK12892 (1-3106)

```
Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 262

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 322

Qy         41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        323 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 382

Qy         61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer  80

Db        383 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 442

Qy         81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100

Db        443 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 502

Qy        101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120

Db        503 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 562

Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140

Db        563 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 622

Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160

Db        623 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 682

Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180

Db        683 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 742

Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200

Db        743 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 802

Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220

Db        803 GATCCAAAATTATGCCCTGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 862

Qy        221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240

Db        863 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 922

Qy        241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260

Db        923 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 982

Qy        261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280

Db        983 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 1042

Qy        281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300

Db       1043 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 1102

Qy        301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320

Db       1103 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 1162

Qy        321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340

Db       1163 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1222

Qy        341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360

Db       1223 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1282

Qy        361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380

Db       1283 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1342

Qy        381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400

Db       1343 CGCTCCCAGACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1402

Qy        401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420

Db       1403 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1462

Qy        421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440

Db       1463 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1522

Qy        441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460

Db       1523 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 1582

Qy        461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480

Db       1583 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1642

Qy        481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500

Db       1643 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1702

Qy        501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520

Db       1703 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1762

Qy        521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540

Db       1763 CTGGTATATTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1822

Qy        541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560

Db       1823 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 1882

Qy        561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580

Db       1883 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 1942

Qy        581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600

Db       1943 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 2002

Qy        601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620

Db       2003 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 2062

Qy        621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640

Db       2063 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 2122

Qy        641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660

Db       2123 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTG 2182

Qy        661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680

Db       2183 GTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGT 2242

Qy        681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700

Db       2243 TATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGC 2302

Qy        701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720

Db       2303 GCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATAT 2362

Qy        721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740

Db       2363 CTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTAT 2422

Qy        741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760

Db       2423 GGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATT 2482

Qy        761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780

Db       2483 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 2542
```

```
Qy        781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2543 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 2602

Qy        801 LysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2603 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 2662

Qy        821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2663 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 2722

Qy        841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2723 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATTAT 2782

Qy        861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2783 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 2842

Qy        881 ValIle 882
              ||||||
Db       2843 GTGATA 2848
```

RESULT 4
AAC85694
ID AAC85694 standard; cDNA; 3120 BP.
XX
AC AAC85694;
XX
DT 29-JUN-2001 (first entry)
XX
DE Nucleotide sequence of human DPP8.
XX
KW Human; dipeptidyl aminopeptidase; DPP8; prolyl oligopeptidase;
KW dipeptidyl peptidase; DPPIV; T cell; cleavage; diarrhoea;
KW growth hormone deficiency; glucose level; mucosal regeneration;
KW non-insulin dependent diabetes mellitus;glucose intolerance;
KW immunosuppression; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 214..2862
FT /*tag= a
FT /product= "Human DPP8"
XX
PN WO200119866-A1.
XX
PD 22-MAR-2001.
XX
PF 11-SEP-2000; 2000WO-AU01085.
XX
PR 10-SEP-1999; 99AU-0002762.
PR 18-FEB-2000; 2000AU-0005709.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Abbott CA, Gorell MD;
XX
DR WPI; 2001-281520/29.
DR P-PSDB; AAB47187.
XX
PT New human dipeptidyl aminopeptidase (DPP8) useful for cleaving
PT substrates, identifying inhibitors of DPP8 catalytic activity which
PT have therapeutic uses, and for detecting activated T cells -
XX
PS Claim 16; Fig 2; 78pp; English.
XX
CC This sequence encodes human dipeptidyl aminopeptidase (DPP8). DPP8
CC has substrate specificity for H-Ala-Pro-pNA, H-Gly-Pro-pNA and
CC H-Arg-Pro-pNA. Therefore, it is a prolyl oligopeptidase and a
CC dipeptidyl peptidase, because it is capable of hydrolysing the
CC peptide bond C-terminal to Pro in each of these compounds. DPP8
CC is homologous with human DPPIV. DPP8 is useful for cleaving a
CC substrate, and for detecting an activated T cell which involves
CC measuring the level of DPP8 gene expression in a T cell. The level
CC of DPP8 expression is detected by detecting the amount of DPP8 RNA
CC in the cell. It is also useful for identifying a molecule capable
CC of inhibiting the cleavage of the substrate by DPP8. Molecules
CC identified as inhibiting DPP8 catalytic activity may be useful for
CC treating diarrhoea, growth hormone deficiency, lowering glucose levels
CC in non-insulin dependent diabetes mellitus and other disorders
CC involving glucose intolerance, enhancing mucosal regeneration and
CC as immunosuppressants.
XX
SQ Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;

Alignment Scores:

Pred. No.:	0	Length:	3120
Score:	4700.00	Matches:	882
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	22	Gaps:	0

US-10-070-464-1 (1-882) x AAC85694 (1-3120)

```
Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 273

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 333

Qy         41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 393

Qy         61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        394 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 453

Qy         81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 513

Qy        101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 573

Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 633

Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 693

Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        694 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 753

Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        754 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 813

Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        814 GATCCAAAATTATGCCCCGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 873

Qy        221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        874 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 933
```

```
Qy       241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
Db       934 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 993

Qy       261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
Db       994 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 1053

Qy       281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
Db       1054 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 1113

Qy       301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
Db       1114 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 1173

Qy       321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
Db       1174 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1233

Qy       341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
Db       1234 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1293

Qy       361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
Db       1294 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1353

Qy       381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
Db       1354 CGCTCCCAGACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1413

Qy       401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
Db       1414 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1473

Qy       421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
Db       1474 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1533

Qy       441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
Db       1534 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 1593

Qy       461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
Db       1594 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1653

Qy       481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
Db       1654 GGGCTGCCTGCTCCAAGTGATTTCAAGTGTCCTATCAAAGAGGAGATAGCAATTACCAGT 1713

Qy       501 GlyGluTrpGluValLeuGlyArgHisGlySerAsnIleGlnValAspGluValArgArg 520
Db       1714 GGTGAATGGGAAGTTCTTGGCCGGCATGGATCTAATATCCAAGTTGATGAAGTCAGAAGG 1773

Qy       521 LeuValTyrPheGluGlyThrLysAspSerProLeuGluHisHisLeuTyrValValSer 540
Db       1774 CTGGTATATTTTGAAGGCACCAAAGACTCCCCTTTAGAGCATCACCTGTACGTAGTCAGT 1833

Qy       541 TyrValAsnProGlyGluValThrArgLeuThrAspArgGlyTyrSerHisSerCysCys 560
Db       1834 TACGTAAATCCTGGAGAGGTGACAAGGCTGACTGACCGTGGCTACTCACATTCTTGCTGC 1893

Qy       561 IleSerGlnHisCysAspPhePheIleSerLysTyrSerAsnGlnLysAsnProHisCys 580
Db       1894 ATCAGTCAGCACTGTGACTTCTTTATAAGTAAGTATAGTAACCAGAAGAATCCACACTGT 1953

Qy       581 ValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLysThrLysGluPhe 600
Db       1954 GTGTCCCTTTACAAGCTATCAAGTCCTGAAGATGACCCAACTTGCAAAACAAAGGAATTT 2013

Qy       601 TrpAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrProProGluIlePhe 620
Db       2014 TGGGCCACCATTTTGGATTCAGCAGGTCCTCTTCCTGACTATACTCCTCCAGAAATTTTC 2073

Qy       621 SerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysProHisAspLeu 640
Db       2074 TCTTTTGAAAGTACTACTGGATTTACATTGTATGGGATGCTCTACAAGCCTCATGATCTA 2133

Qy       641 GlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyGlyProGlnValGlnLeu 660
Db       2134 CAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATATATGGTGGTCCTCAGGTGCAGTTG 2193

Qy       661 ValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAlaSerLeuGly 680
Db       2194 GTGAATAATCGGTTTAAAGGAGTCAAGTATTTCCGCTTGAATACCCTAGCCTCTCTAGGT 2253

Qy       681 TyrValValValValIleAspAsnArgGlySerCysHisArgGlyLeuLysPheGluGly 700
Db       2254 TATGTGGTTGTAGTGATAGACAACAGGGGATCCTGTCACCGAGGGCTTAAATTTGAAGGC 2313

Qy       701 AlaPheLysTyrLysMetGlyGlnIleGluIleAspAspGlnValGluGlyLeuGlnTyr 720
Db       2314 GCCTTTAAATATAAAATGGGTCAAATAGAAATTGACGATCAGGTGGAAGGACTCCAATAT 2373

Qy       721 LeuAlaSerArgTyrAspPheIleAspLeuAspArgValGlyIleHisGlyTrpSerTyr 740
Db       2374 CTAGCTTCTCGATATGATTTCATTGACTTAGATCGTGTGGGCATCCACGGCTGGTCCTAT 2433

Qy       741 GlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIle 760
Db       2434 GGAGGATACCTCTCCCTGATGGCATTAATGCAGAGGTCAGATATCTTCAGGGTTGCTATT 2493

Qy       761 AlaGlyAlaProValThrLeuTrpIlePheTyrAspThrGlyTyrThrGluArgTyrMet 780
Db       2494 GCTGGGGCCCCAGTCACTCTGTGGATCTTCTATGATACAGGATACACGGAACGTTATATG 2553

Qy       781 GlyHisProAspGlnAsnGluGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlu 800
Db       2554 GGTCACCCTGACCAGAATGAACAGGGCTATTACTTAGGATCTGTGGCCATGCAAGCAGAA 2613

Qy       801 LysProProSerGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 820
Db       2614 AAGTTCCCCTCTGAACCAAATCGTTTACTGCTCTTACATGGTTTCCTGGATGAGAATGTC 2673

Qy       821 HisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLysProTyrAsp 840
Db       2674 CATTTTGCACATACCAGTATATTACTGAGTTTTTTAGTGAGGGCTGGAAAGCCATATGAT 2733

Qy       841 LeuGlnIleTyrProGlnGluArgHisSerIleArgValProGluSerGlyGluHisTyr 860
Db       2734 TTACAGATCTATCCTCAGGAGAGACACAGCATAAGAGTTCCTGAATCGGGAGAACATTAT 2793

Qy       861 GluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeuGlySerArgIleAlaAlaLeuLys 880
Db       2794 GAACTGCATCTTTTGCACTACCTTCAAGAAAACCTTGGATCACGTATTGCTGCTCTAAAA 2853

Qy       881 ValIle 882
             ||||||
Db       2854 GTGATA 2859
```

RESULT 5
AAD38956

ID AAD38956 standard; cDNA; 3120 BP.
XX
AC AAD38956;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human dipeptidyl peptidase 8 (DPP8) cDNA.
XX
KW Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
KW autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;
KW graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
KW antiviral; enzyme; gene; ss.

```
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             214..2862
FT                   /*tag=  a
FT                   /product= "Human DPP8 protein"
XX
PN   WO200234900-A1.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-AU01388.
XX
PR   27-OCT-2000; 2000AU-0001078.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Abbott CA,  Gorrell MD;
XX
DR   WPI; 2002-454646/48.
DR   P-PSDB; AAE24170.
XX
PT   New dipeptidyl peptidase (DPP) peptides, useful for screening
PT   inhibitors of DPP catalytic activity, which may be employed to treat
PT   e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PT   rejection and HIV infection -
XX
PS   Example; Fig 1; 91pp; English.
XX
CC   The present invention relates to dipeptidyl peptidase (DPP) proteins and
CC   polynucleotides encoding such proteins. The DPP peptides are useful for
CC   screening inhibitors of DPP catalytic activity. The inhibitors are useful
CC   for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
CC   rejection and HIV (human immuno deficiency virus) infection. The present
CC   sequence is human DPP8 cDNA.
XX
SQ   Sequence 3120 BP; 936 A; 637 C; 706 G; 841 T; 0 other;

Alignment Scores:
Pred. No.:                0              Length:          3120
Score:                    4700.00        Matches:         882
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       24             Gaps:            0

US-10-070-464-1 (1-882) x AAD38956 (1-3120)

Qy          1 MetAlaAlaAlaMetGluThrGluGlnLeuGlyValGluIlePheGluThrAlaAspCys 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 ATGGCAGCAGCAATGGAAACAGAACAGCTGGGTGTTGAGATATTTGAAACTGCGGACTGT 273

Qy         21 GluGluAsnIleGluSerGlnAspArgProLysLeuGluProPheTyrValGluArgTyr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 GAGGAGAATATTGAATCACAGGATCGGCCTAAATTGGAGCCTTTTTATGTTGAGCGGTAT 333

Qy         41 SerTrpSerGlnLeuLysLysLeuLeuAlaAspThrArgLysTyrHisGlyTyrMetMet 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 TCCTGGAGTCAGCTTAAAAAGCTGCTTGCCGATACCAGAAAATATCATGGCTACATGATG 393

Qy         61 AlaLysAlaProHisAspPheMetPheValLysArgAsnAspProAspGlyProHisSer 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        394 GCTAAGGCACCACATGATTTCATGTTTGTGAAGAGGAATGATCCAGATGGACCTCATTCA 453

Qy         81 AspArgIleTyrTyrLeuAlaMetSerGlyGluAsnArgGluAsnThrLeuPheTyrSer 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        454 GACAGAATCTATTACCTTGCCATGTCTGGTGAGAACAGAGAAAATACACTGTTTTATTCT 513

Qy        101 GluIleProLysThrIleAsnArgAlaAlaValLeuMetLeuSerTrpLysProLeuLeu 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        514 GAAATTCCCAAAACTATCAATAGAGCAGCAGTCTTAATGCTCTCTTGGAAGCCTCTTTTG 573

Qy        121 AspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGluGluGluLeuLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 GATCTTTTTCAGGCAACACTGGACTATGGAATGTATTCTCGAGAAGAAGAACTATTAAGA 633

Qy        141 GluArgLysArgIleGlyThrValGlyIleAlaSerTyrAspTyrHisGlnGlySerGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 GAAAGAAAACGCATTGGAACAGTCGGAATTGCTTCTTACGATTATCACCAAGGAAGTGGA 693

Qy        161 ThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLysAspGlyGlyProGlnGly 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        694 ACATTTCTGTTTCAAGCCGGTAGTGGAATTTATCACGTAAAAGATGGAGGGCCACAAGGA 753

Qy        181 PheThrGlnGlnProLeuArgProAsnLeuValGluThrSerCysProAsnIleArgMet 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        754 TTTACGCAACAACCTTTAAGGCCCAATCTAGTGGAAACTAGTTGTCCCAACATACGGATG 813

Qy        201 AspProLysLeuCysProAlaAspProAspTrpIleAlaPheIleHisSerAsnAspIle 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        814 GATCCAAAATTATGCCCCGCTGATCCAGACTGGATTGCTTTTATACATAGCAACGATATT 873

Qy        221 TrpIleSerAsnIleValThrArgGluGluArgArgLeuThrTyrValHisAsnGluLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        874 TGGATATCTAACATCGTAACCAGAGAAGAAAGGAGACTCACTTATGTGCACAATGAGCTA 933

Qy        241 AlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaThrPheValLeuGlnGluGlu 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        934 GCCAACATGGAAGAAGATGCCAGATCAGCTGGAGTCGCTACCTTTGTTCTCCAAGAAGAA 993

Qy        261 PheAspArgTyrSerGlyTyrTrpTrpCysProLysAlaGluThrThrProSerGlyGly 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        994 TTTGATAGATATTCTGGCTATTGGTGGTGTCCAAAAGCTGAAACAACTCCCAGTGGTGGT 1053

Qy        281 LysIleLeuArgIleLeuTyrGluGluAsnAspGluSerGluValGluIleIleHisVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1054 AAAATTCTTAGAATTCTATATGAAGAAAATGATGAATCTGAGGTGGAAATTATTCATGTT 1113

Qy        301 ThrSerProMetLeuGluThrArgArgAlaAspSerPheArgTyrProLysThrGlyThr 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 ACATCCCCTATGTTGGAAACAAGGAGGGCAGATTCATTCCGTTATCCTAAAACAGGTACA 1173

Qy        321 AlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGlyArgIle 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1174 GCAAATCCTAAAGTCACTTTTAAGATGTCAGAAATAATGATTGATGCTGAAGGAAGGATC 1233

Qy        341 IleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGlyValGlu 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1234 ATAGATGTCATAGATAAGGAACTAATTCAACCTTTTGAGATTCTATTTGAAGGAGTTGAA 1293

Qy        361 TyrIleAlaArgAlaGlyTrpThrProGluGlyLysTyrAlaTrpSerIleLeuLeuAsp 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1294 TATATTGCCAGAGCTGGATGGACTCCTGAGGGAAAATATGCTTGGTCCATCCTACTAGAT 1353

Qy        381 ArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIleProValGlu 400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1354 CGCTCCCAGACTCGCCTACAGATAGTGTTGATCTCACCTGAATTATTTATCCCAGTAGAA 1413

Qy        401 AspAspValMetGluArgGlnArgLeuIleGluSerValProAspSerValThrProLeu 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1414 GATGATGTTATGGAAAGGCAGAGACTCATTGAGTCAGTGCCTGATTCTGTGACGCCACTA 1473

Qy        421 IleIleTyrGluGluThrThrAspIleTrpIleAsnIleHisAspIlePheHisValPhe 440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1474 ATTATCTATGAAGAAACAACAGACATCTGGATAAATATCCATGACATCTTTCATGTTTTT 1533

Qy        441 ProGlnSerHisGluGluGluIleGluPheIlePheAlaSerGluCysLysThrGlyPhe 460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1534 CCCCAAAGTCACGAAGAGGAAATTGAGTTTATTTTTGCCTCTGAATGCAAAACAGGTTTC 1593

Qy        461 ArgHisLeuTyrLysIleThrSerIleLeuLysGluSerLysTyrLysArgSerSerGly 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1594 CGTCATTTATACAAAATTACATCTATTTTAAAGGAAAGCAAATATAAACGATCCAGTGGT 1653

Qy        481 GlyLeuProAlaProSerAspPheLysCysProIleLysGluGluIleAlaIleThrSer 500
```